

DOCUMENT RESUME

ED 295 330

EA 020 094

AUTHOR Hannan, Bill
TITLE Assessment and Evaluation in Schooling. ESA844, Administrative Context of Schooling.
INSTITUTION Deakin Univ., Victoria (Australia).
REPORT NO ISEN-0-7300-0300-0
PUB DATE 85
NOTE 144p.; For related documents, see EA 020 080-081 and EA 020 091-093.
AVAILABLE FROM Publication Sales, Deakin University Press, Deakin University, Geelong, Victoria 3217, Australia (Stock No.ESA844M03; \$12.50 Australian; quantity and educational discounts).
PUB TYPE Viewpoints (120) -- Collected Works - General (020) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Educational Assessment; *Educational Philosophy; Elementary Secondary Education; Evaluation Criteria; Foreign Countries; *Foundations of Education; *Grading; *Marxian Analysis; Political Socialization; Social Structure; Social Values; *Student Evaluation
IDENTIFIERS Australia (Victoria)

ABSTRACT

This volume is part of a series of monographs from Australia devoted to outlining an alternative approach, based on neo-Marxist concepts, to educational administration. The opening discussion is a critique of conventional student evaluation practices aimed at tracing the social influences that are brought to bear on students, teachers, schools, families, and knowledge itself. The first section argues that educational institutions use a variety of concepts of "ability" and "educability" that are adopted more for their political import or their administrative convenience than for their closeness to scientifically established fact. Remaining sections focus on the following topics: pre- and post-assessment, competitive and noncompetitive assessment, the movement against grading, and the four styles of assessment: graded (norm-referenced), descriptive (individual-referenced), mastery (criterion-referenced), and work-based (goal-referenced). The second part of the paper moves from student assessment per se to evaluation of the social values underpinning each style and the contribution of each to maintaining intellectual standards, classroom management, student socialization, and pedagogy. The conclusion calls for a shift of emphasis from assessment of students to evaluation of courses. Five readings by separate authors follow: (1) "Educational versus Managerial Evaluation in Schools," by Richard J. Bates; (2) "Education in Theory, Schooling in Practice: The Experience of the Last Hundred Years," by Brian Simon; (3) "Assessment: A Workable Democratic Method," by Bill Hannan; (4) "Assessment in Practice: Competitive or Non-Competitive?" by Graeme Withers and Greg Cornish; and "Year 12," by the Victorian Secondary Teachers Association. An annotated bibliography is included. (TE)

ESA844 Administrative Context of Schooling

Assessment and Evaluation in Schooling

Bill Hannan

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

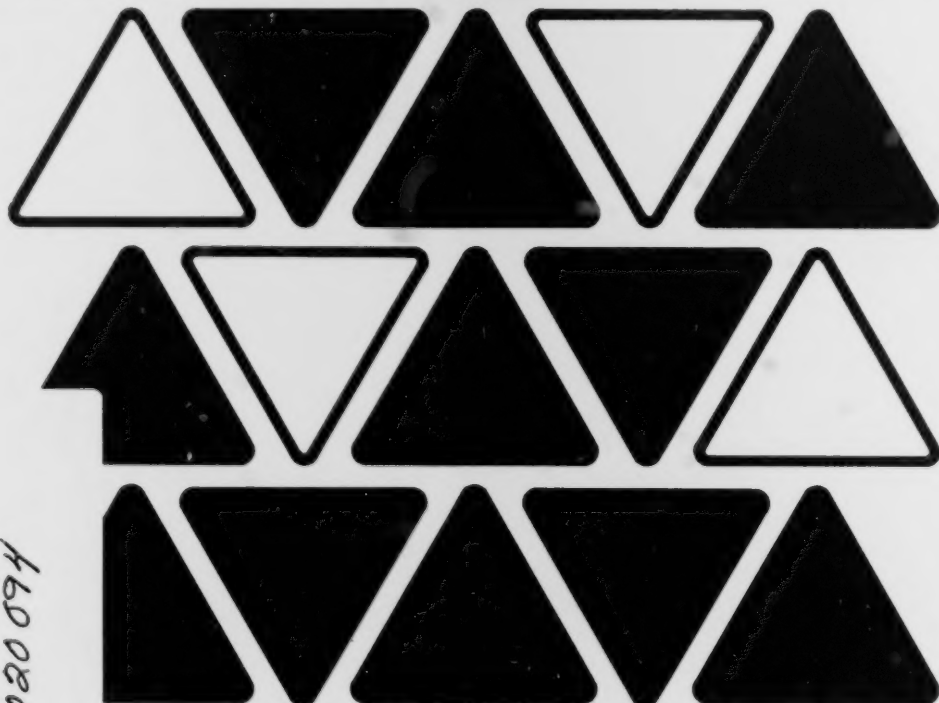
☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL IN MICROFICHE ONLY
HAS BEEN GRANTED BY

Fay Waring

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."



ESA844 Administrative Context of Schooling

Assessment and Evaluation in Schooling

Bill Hannan

Victorian Secondary Teachers Association



Deakin University
Victoria 3217

Published by Deakin University, Victoria 3217
Distributed by Deakin University Press
First published 1985

© Deakin University 1985

Edited and designed by Deakin University Production Unit
Printed by Deakin University Printery

National Library of Australia
Cataloguing-in-publication data

Hannan, W. (William), 1932-
Assessment and evaluation in schooling

Bibliography.

ESA844

ISBN 0 7300 0300 0

1. Educational tests and measurements. 2. Students. Rating of. I. Deakin University School of Education. Open Campus Program. II. Title. III. Title. Administrative context of schooling.

371.2.64

This book forms part of the ESA844 *Administrative Context of Schooling* course offered by the School of Education in Deakin University's Open Campus Program. It has been prepared by Bill Hannan for the ESA844 *Administrative Context of Schooling* course team, whose members are:

Course team

Lawrence Angus
Richard Bates (chairperson)
Diana Macmillan (course developer)
Fazal Rizvi
Peter Watkins

Consultants

Peter Musgrave, Monash University
Bill Hannan, Victorian Secondary Teachers Association

The course includes:

Richard Bates, *Liberalism, Marxism and the Struggle for the State: Prolegomena to the Study of Public Administration*

Richard Bates, *Public Administration and the Crisis of the State*

Bill Hannan, *Assessment and Evaluation in Schooling*

P.W. Musgrave, *Curricular Decisions in their Administrative Contexts*

Peter Watkins, *Agency and Structure: Dialectics in the Administration of Education*

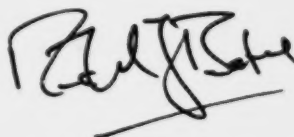
These books are available from Deakin University Press, Deakin University, Victoria 3217

Students enrolled in the course are also supplied with a guide to the course

Series introduction

It is not possible to understand the nature of educational administration without understanding the broader context of public administration or, further, the social and political debate over the nature of the state, civil society and the economy, and their relationships. The series of volumes of which this book is a part addresses these various issues. Beginning with a discussion of the contested relationship between the individual and the state, the politics of administration is set within the debate over liberalism, Marxism and critical theory, and the nature of the crisis of the modern state. The impact of this crisis on public administration is then examined, especially in terms of the 'new' public administration and the notion of public good. An examination of educational administration follows, as do studies of the administrative context of curriculum and of evaluation. Finally, a discussion of the dialectical nature of educational administration is presented.

The introductory essay of each volume is a digest of current debate and a contribution to it. So that readers may enter that debate rapidly key readings are appended, as is an annotated bibliography of key works in the field. We hope that this presentation of the debate will encourage others to join in the exploration of such issues in educational administration.

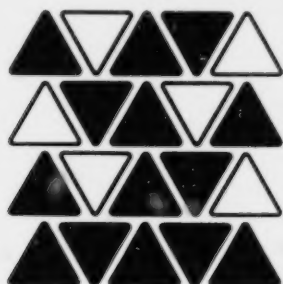


Richard J. Bates
Course team chairperson

Contents

Series introduction	iii
Assessment and evaluation in schooling	1
Introduction	3
Ability and educability	4
Pre- and post-assessment	6
Competitive and non-competitive assessment	7
The movement against grading	9
Styles of assessment	13
From assessment to evaluation	15
Evaluating courses	22
References	24
Readings	25
1 Educational versus managerial evaluation in schools Richard J. Bates	27
2 Education in theory, schooling in practice: The experience of the last hundred years Brian Simon	44
3 Assessment: A workable democratic method Bill Hannan	63
4 Assessment in practice: Competitive or non-competitive? Graeme Withers & Greg Cornish	72
5 Year 12 Victorian Secondary Teachers Association	108
Annotated bibliography	137
About the author	140
Acknowledgements	141

Assessment and evaluation in schooling



Introduction

A school report that gives a student A or B for effort and D or E for achievement encapsulates most of the problems of assessment and evaluation in education. Faced with such a report teachers, students and their families all have a very real question: What do we do next? Hard work, it seems, is not enough. What, then? The written comment on the report will probably confirm rather than resolve the dilemma: 'Mary has worked hard but she finds this subject difficult'.

In practice Mary will repeat the subject, or drop it, or get deeper into the bog in the next stage of it. Probably these will be her only real options. In theory, however, she could challenge the assumption at work, namely that the student is not up to the course. She could well say, possibly with justice, that the course itself was poorly designed, or that it was badly taught. In her terms, then, assessment and evaluation would involve not just the assessment of students' performances but also a critical examination of the content of the course and of the way it is taught.

Taken singly, none of these dimensions will necessarily throw light on what has happened to Mary. In fact, a process is going on in which both student and teacher are caught quite centrally, and for which responsibility spreads throughout the community to families, to schools, to systems and finally to the complex of cultural agents which produce knowledge in the form in which we use it. An evaluation of such a large and diffuse process is necessarily in itself a shifting and dynamic process. It is a question of tracing the influences at work on students, teachers, schools, families and on knowledge itself, and of trying to juggle them into a coherent pattern.

If this is true, and it is not said merely to be daunting but in order to get to some positive ends, then the common practice of blaming the student for the failure of the process is among the crudest of possible evaluations, a sort of payment by results exacted from the students. Yet it is the most common form of evaluation practised in Australian schools. And we should be clear that it is practised right through the schooling process. Student assessment and examination results, which are invariably seen as measures of student ability, are only the tip of the iceberg. Beneath those apparently secure and timeless waters lies a mountain range of historical practices built on the supposition that most elements of the process can be regarded as fixed and that student ability is therefore the critical variable.

In order to avoid that crude view, it is as well to ascribe different meanings to our two terms, 'assessment' and 'evaluation'. I suggest that we might reserve 'assessment' to refer to judgements about students' work. 'Evaluation' then would refer to judgements about courses and pedagogy. In this monograph, I will use the terms that way, and one of my chief contentions will be that the two are interdependent, and in particular, that assessment of students should not take place outside the evaluation of courses and pedagogy.

I should also make it clear that I am writing about those phases of schooling that can or should be regarded as universal. This of course includes

all of the compulsory period, which is certainly the essential phase on which to focus our discussion. There is, however, an argument to be made for including the final two years of secondary schooling and whatever is used as the transition to tertiary or further education. In part, this is a numbers argument. Over eighty per cent of students stay on into Year 11, and there are still about sixty per cent in some form of schooling or training by Year 12. But, more importantly, it is a cultural argument. People should be encouraged to stay on because education is useful in our kind of society. It is useful to individuals both in their work and in their private and social lives; and it is useful to the community because it helps people to be more productive and more civilised.

Of course, whether or not education does in fact accomplish these desirable ends will depend not only on its own quality but also on what wider economic and social policies will allow it to achieve. Generally, however, I would maintain that an inclusive and responsive schooling is one of the necessary bases for achieving those ends; and I would also argue that our society probably needs to guarantee access to twelve years of schooling and that a period of post-secondary training of, say, two or three years should be available to those wanting it either immediately after schooling or at some later stage. Of course, by the time one is at the upper secondary or tertiary level, significant degrees of specialisation go along with 'general' elements. None the less, many of the cultural arguments that must apply fully to compulsory education should be seen to apply significantly also to post-compulsory schooling as it is now developing.

Ability and educability

Before we trace the practices of assessment and evaluation as they now appear in the school systems, we should consider very briefly the idea of ability. I say briefly because I am not a psychologist or a learning theorist. My ideas about ability are those one has to form in the course of being a teacher, a parent and a writer about schools. From that vantage point two features stand out: one is that the concept of ability is not settled, but highly controversial; and the other is that none of the contending concepts of ability indicates clearly how we might organise schools that will cater for everyone.

The result is that educational institutions use a variety of concepts that are adopted more for their political import or their administrative convenience than for their closeness to scientifically established fact. Thus, those wanting to be selective — universities and up-market schools, for example — are much more likely to believe that natural intelligence can be measured and taught to, than are institutions that are consciously 'open'. In fact, the selective and the open institution represent two landmarks on opposite horizons in the history of education, the former with largely deterministic ideas about inherited ability and the latter with optimistic ideas about human educability. These are, of course, extreme characterisations. In real life it is conceded that 'intelligence' (in many of its forms) may also be acquired,

and that there are considerable limits in practice to many people's educability.

The problem with this state of affairs is the very centrality of the concept. It is a bit like the Afterlife: if it exists, it is very important. If there is such a faculty as intelligence — a kind of generalised ability to perform intellectual work — if it could be identified and if intellectual work could be efficiently matched to it, then a certain pattern of education and institutions could be imagined. If, however, one questions how generalisable the concept of intelligence is, whether it can be identified, and especially whether it is an inherited attribute rather than a cultural construct, then another, radically different education would result. And, of course, one's approach to matters such as assessment and evaluation would be qualitatively different. Historically, we have examples of education systems based on elements of each of those extremes. In Australia, with the steady expansion of secondary and, to a lesser extent, tertiary education we have seen the two extremes mixed and watched them conflict. On the one hand, pushed by demand, we have pursued the ideal of educability by admitting more and more students to higher levels of education, and on the other we have persisted with a tradition of stratified schooling with our academic high schools, both state and private, and our district high, technical and comprehensive schools. Accompanying that has been a parallel conflict in curriculum between general and instrumental education, the former seeking a more sophisticated, intellectual and cultural formation for all, the latter a differentiated matching of courses and standards to the labour market.

The fact that these two ideas or the various shades of them continue to conflict within our education system is evidence in part that neither has a firm basis in theory. Rather, they are ideas that can command both some learned and some popular support. Intellectually, they are open questions. It is useless to deny the evidence of common observation that people are good or not good at various things; but against that we can put equally familiar observations of the power of the human mind to master highly complex systems and ideas — language, for example. Whilst it is of intense interest and importance for us, education or not, to discover what goes into making these abilities, the immediate question to decide is how important the issue is to basic schooling, say up to Year 11 or 12. In the absence of clearcut knowledge, this decision becomes a pragmatic and a politically significant one.

In the context of education during the past half-century the more convenient decision has been to endorse 'ability' rather than 'educability'; and the consequence of that has been to create and reinforce a popular belief that children have a more or less fixed capacity to learn, and that this capacity varies widely even in very young children. This belief is then called upon to sustain practices which spread students out by statistical methods — 'normal curves of distribution', and the like — so that even the most basic activities of schooling, such as reading, writing and counting, have their good, average and bad students. Such a belief focuses on differences, such as pace, rather than on common capacities. Thus, the speed with which

various members of a class learn to read can be made to count more than the fact that virtually all can learn to read. And that emphasis on pace can become not only the most salient feature of students' assessments but also a decisive factor in the grouping of students and in the construction of the materials from which they learn.

Why has that emphasis on differentiated abilities been the more convenient to follow? At least two reasons suggest themselves. The first is the intellectual influence, experienced particularly through teacher training and through systemic practices such as student placement, of the schools of educational psychology that have pursued mental testing. The IQ is the best known form of this preoccupation with measuring assumed ability, but in current school practice the more common form is probably the reading and number tests used to place primary and junior secondary students in groups, classes or schools. The second reason is to be found in the greatly increased use of schools as the arena for social competition. Competition requires a systematic selection of participants and a scarcity value in the rewards. In Australia this is in fact largely focused on certain tertiary and occupational qualifications, but one of the pressure points of this process is the credential which gives access to those qualifications—the Higher School Certificate (HSC). In most states, the HSC is norm-referenced, and this practice tends to flow down to earlier years. Thus the notion of the statistical distribution of ability is massively reinforced. The very statistical symmetry of the process lends a scientific air to the assumptions that must be made in order to operate the process. Yet they are assumptions of the most dubious kind: largely to the effect that students' performances in extremely different areas of knowledge can be measured quantitatively, compared and aggregated to yield a finely scaled rank order. None the less, the social pressure towards competition is so strong that anything that lends an air of fairness and scientific objectivity to the selection process is readily seized upon.

Beyond expedience, and perhaps habit, there seems to be no justification for taking an extreme view of ability. In the absence of certain knowledge, it would seem more prudent for public education policy to be premised on the educability of the great majority during the phase of schooling held to be necessary for an adequate cultural formation.

At all events, whatever decision is made, it should be clear that it will have truly profound effects on the nature and outcomes of schooling. In particular, all aspects of educational assessment and evaluation will be permeated and coloured by the participants' attitudes to human abilities and capacities.

We should now look at some of those aspects in a little more detail in order to illustrate just how deeply the practices associated with different kinds of assessment and evaluation penetrate the layers of schooling.

Pre- and post-assessment

As I said at the beginning, student assessments, in the form of school reports, are the most visible point of the assessment and evaluation process. In turn,

however, they are themselves built on a series of judgements that may spread so widely that they appear to be unconnected with the assessment. This unconnectedness is probably most characteristic of normative assessment. Since a normative assessment appears to relate a student's performance fairly directly to a student's capacity, the fact that this same idea of capacity has already greatly influenced the student's learning may be hidden. Yet it will probably have been used to place the student in a particular group or class or even school to begin with. The state systems in Australia all tolerate or operate elite schools which select their intakes academically. The same states still implicitly tolerate, at the middle and upper secondary levels, the distinction between 'academic' and 'comprehensive' or 'technical' schooling. And all either advocate or tolerate various versions of streaming. So in the very formation of learning groups, it is more likely than not that a process of student assessment has taken place. If students are coming into what is seen as a distinct phase — say, from infants to primary or from primary to secondary — they are likely to be formally tested either for general ability or on their basic skills, again in a general way. Assessment, in other words, is not exercised solely after the fact — forms of pre-assessment are also involved.

From this double-barrelled assessment very different courses may be derived. Low streams will certainly be seen as going more slowly than others; but it is likely that they will also cover a different content — not simply less of the same, or the same less deeply, but a different content held to be more relevant or useful to them. In other words, there is more at stake than differences of pace or skill. Cultural differences may be opened up that perhaps need not have existed, and real differences may be deepened to the point where they become unbridgeable.

Once these same differentiating assumptions are applied also to teaching materials and methods, we can arrive at a virtually closed system, locked up by its own assumptions. Teachers brought under pressure to change (to a non-competitive system, for example) can genuinely see no alternative to their classroom organisation or teaching materials. Teachers wanting to change find themselves having to recast their whole course, write new materials and answer criticism of their methods. This last, of course, is no bad thing, but it is something that the teacher of a competitive classroom normally does not have to do. Much, in other words, will follow for the teacher who decides to stop grading students from A to E.

Competitive and non-competitive assessment

What makes assessment such a lively issue is that many teachers and schools have in fact tried to reform entrenched methods (see Hannan 1982). The educational questions involved in both assessment and evaluation have been forged in the fires of controversy and put to the test by contending practices. In Victoria at least, this contention has centred on the place of competition in schooling. Historically, it is argued, school has increasingly become the site for social competition. When, in the first half of the century, elementary school was still the only level open to the masses, competition

for the wealth or prestige associated with education was confined to a meritorious elite drawn from the poor to be schooled alongside those who by birth were entitled to more than elementary schooling. From the 1950s however, this elite expanded to a mass, and schooling became a major focus — arguably the major focus — of social competition. This competition was channelled by the credential — the HSC — which served largely to select students for entry to prestigious and lucrative professions. Thus upper secondary schooling developed into a process for ranking and selecting students, and the rest of secondary schooling tended more and more to be treated as a run-up to the credential. A 'good' school became one which was choosy with its intake and, as a result, delivered good results in the HSC. Schools who took in all comers slipped down the ladder, if indeed they had ever had a foot on it.

This competition for the very real social rewards of schooling found some ready-made vehicles in existing school practices. Periodical testing and examining are the most obvious of these, but the roles of streaming and tracking within a school and of the hierarchy of private, elite state, comprehensive and technical schools should not be overlooked. To make these practices even more effective (or apparently so) educational psychology and measurement provided increasingly sophisticated statistical methods of testing and comparing students' performances. These methods are frequently challenged (see, for example, Houts 1977). They are said to have the effect of narrowing what is taught to what may be measured, or alternatively of measuring crudely and inaccurately things that are not really measurable. The methods of comparing performances among students are even more open to criticism especially when they attempt to establish some aggregated index of students' performances across a number of subjects. What real validity is there in attempting to weight and total marks gained in, say, a written examination in geography, a set of problems in mathematics, an oral test in French and practical work in art? Further, what account, if any, should be taken of students' backgrounds, of the way they have been taught, and so on? Despite such criticisms, however, the manipulation of students' performances according to various statistical models has become entrenched, probably because it gives an air of precise mathematical objectivity, and hence fairness, to judgements upon which a great deal may depend.

The sophisticated juggling of marks in public examinations such as the HSC is not of course carried down into lower levels of school, or at least not to the same degree. At the same time, however, the trickle-down effect is substantial. Although teachers may vary a lot in their practices at any one time, some passing most of their class, others failing large numbers, there is a tendency over time for the school population to settle into a pyramidal shape with a mounting number of failures and a diminishing proportion of successes year by year.

When a pyramid such as this is constructed, with the majority of students progressively reaching some kind of ceiling and either dropping out or taking

up 'alternative' tracks, we are faced with a special kind of competition, one that has built-in failure as well as success.

To get far in public debate, you do need to make some distinctions about the meaning of competition. In my experience, very few people think of competition unfavourably. For most it is an incentive, necessary to get the best out of people. For many it is an integral and desirable fact of modern life. For teachers, with large groups to manage, competition can be a constructive and convenient form of incentive and control. There is not much to be gained by disputing that. The point to question is whether it is used as an incentive for all, or whether, in practice, it is an incentive for the regular winners and an actual disincentive for the regular failures. The pyramid, with its built-in ceilings and necessary proportion of failures, is of the second kind and is in fact a destructive form of competition.

It is this form of competition, based on assumptions about the way abilities are distributed throughout the population, which has been attacked over the past twenty years in Victoria (Hannan, in press). The specific target has been grading, in particular the five-point letter scale from A to E. In its place, descriptive assessments were first proposed. Thus a group of eighteen teachers from Preston Girls High School, who were fined in 1980 for refusing to use grades, described their form of assessment as 'non-competitive descriptive assessment' (Grant and Hannan 1981).

The phrase is telling, because it not only says what the assessment is — 'descriptive' — but throws out a challenge by declaring what it is not. It is consciously against the prevailing order. It is out to undermine the pyramid.

An account of this movement to create a non-competitive form of assessment reaching even into the credential will illustrate both the issues and the forces that have to be confronted.

The movement against grading

To my knowledge, active opposition to grading began in Victorian schools in 1965 (Hannan, in press). By then, the radical non-competitive ethos of John McLeod, head of the secondary section of the Victorian Education Department's Psychology and Guidance Branch, was becoming known to teachers and being put into practice by David Schapper, founding principal of Maryvale High School. Also in that year some English teachers (including myself) at Moreland High School refused to grade the work being done in their junior classes. The school's principal at first resisted this, but eventually he conceded that, even if we were not right, we were determined and honestly convinced that our students would be better off, and backed us in altering the form of reporting and explaining it to parents.

It is not entirely fair to claim this as the first move away from grading. In fact, some subjects such as art, physical education and religious education had long been ambivalent about giving marks or grades. Art teachers, for instance, who saw their aim as drawing on the creativity they believed existed in every child, found it contradictory to measure and grade such

an elusive quality. Physical education teachers had a similar commitment to bringing out the potential in all, because they saw physical well-being as something necessary to all people throughout life rather than as something reserved for the unusually swift, strong or agile. A similar moral dilemma obviously dogs religious education.

These subjects, however, were (and still are) seen as peripheral to the core academic subjects. English, by contrast, is the core of the core. So when we at Moreland applied to the teaching of English arguments about creativity, life-long needs and universal success we were rocking the boat. If such arguments could be sustained for English it is possible that they could be sustained for the whole curriculum and be applied more widely to selective school practices such as streaming and tracking.

How much ground these arguments really gained over the next decade has not, to my knowledge, been comprehensively surveyed. Between 1966 and 1970, non-competitive assessment became part of much more general moves to reform the curriculum of secondary schools. These moves were started by R. A. Reed, then Victorian Director of Secondary Education, who in 1966 set up a representative Curriculum Advisory Board (CAB) which over the next few years produced a statement of principles, and sponsored a number of papers and some pilot schools, one of which was Moreland High School. Reed asserted that secondary schooling should be a six-year phase of education in its own right – free, that is, from external control through the credential – and included as one of the principles of the CAB the following statement:

There is no place for competitive assessment in Secondary School. Whatever assessment is done should be seen as a function of the essential communication between school and child and between school and parents.

(Hannan, *in press*)

It is safe to read this as a proscription of the kind of competitive assessment practices that build in failure, of normative assessment in fact. It may also be taken as an attack on the effects of the HSC at the top end of schooling and of various kinds of mental testing at the bottom end. The influence of John McLeod would have been strong in this. McLeod was a radical social psychologist who saw social circumstances as the key to learning. He rejected the deterministic assumptions on which intelligence testing and general ability testing were based, and he denied that there was any adequate psychological theory of learning on which teaching could be based. (To my knowledge, he never wrote any of this down, but he said it to a lot of influential people.) His position in the Psychology and Guidance Branch of the Victorian Education Department allowed him to keep IQ testing out of the schools and his courses in school-based counselling extended his intellectual influence quite markedly in the mid- to late 1960s. Other members of the Curriculum Advisory Board elaborated this line of thought to include opposition not only to grading and selective certification but also to streaming, tracking and other forms of stratification within individual schools or within the whole system (Dow 1970, p.21).

The context in which this was happening, however, is all important. As the next decade was to show there are two quite distinct paths following from the notion of non-competitiveness in curriculum—one which tries to take everyone through a common enterprise and one which takes its direction from individual differences. I believe that the first of these would have been in most people's minds in 1968. Despite its streams and its tracks, the curriculum was still generally common. Basically, it was the academic curriculum of more elite days adapted, through streaming, to differences of pace or depth and modified from Year 9, through tracking, to give those who would leave at the age of 14 some marketable or social skills, namely commercial and domestic work for girls and some trade background for boys. New South Wales's Wyndham Scheme was the most complete bureaucratic expression of this curriculum. Its ideal was the Scottish 'lad o' pairts', the working class or country boy who could get himself a useful general education out of the right sort of public institution. To this ideal student, being in the bottom band for one subject was no barrier to being in the top for another, and Wyndham clearly believed that this would be the usual pattern; all students would have successes to buoy them up and the grit to improve their weak points. A multi-level curriculum would give them that chance, but the outcome would be a common cultural formation for all. Perhaps if that view of human nature, combined with Presbyterian moral force, had reigned in all schools, it would have worked.

Much of the other work supported by the Curriculum Advisory Board illustrates the way in which the idea of non-competitive classrooms was meant to be interpreted. In particular, Board documents deal at some length with the integration of the main academic studies—English, mathematics, science, history, geography, art—and with the freeing of students' working patterns from the strict 'forty students for forty minutes forty times a week' of the timetable (*Secondary Teacher*, 1970). The way to the non-competitive curriculum was seen to lie in a reinterpretation of the traditional academic subjects and a shift to a communicative classroom based on small groups. Discussion, the search for general meanings and the supporting of individuals by the group were essential to the pedagogy. The ideal was a community of scholars, motivated by the satisfaction that comes from understanding.

It is not, of course, possible to measure the penetration of such ideals into school practice. One has impressions, and mine is that the notion of commonality fell on barren soil, whilst that of individual difference took deeper root and suckered in many directions. Throughout the 1970s the idea that non-competitive, descriptive assessments simply recorded a student's progress against previous individual achievement became standard and this reflected a pedagogy in which individuals worked on assignments 'at their own pace' and, in theory at least, each individual could be working not only on a different stage of the same subject-matter but on different subject-matter. At the very least, descriptive assessment accommodated and legitimised this diversification, but it probably also released it.

Theoretically, the only limits to endless diversification to match the presumed or expressed 'needs' of individuals were the classic ones of schooling: class management and instructional materials. And teaching machines and programmed instruction seemed to offer the final solution to those problems.

By the late 1970s, what had started as a progressive reform was ready to be absorbed into a new conservative accommodation to the pressure of mass participation in schooling. School systems were encouraged to diversify and to create alternative paths to suit different groups of students. Choice among these alternatives was then to be exercised by individuals. Conservative theorists depicted this as a valid application of market principles to education: good alternatives would survive, bad ones would go under. (Conservative politicians in England, USA and Australia expressed this theory in the form of proposals for voucher systems. See 'The new liberalism' in the *Weekend Australian*, 23 June, 1984.)

This development is best illustrated by a short account of the agitation against the HSC in Victoria. In some ways the HSC stands as a formidable barrier to diversification despite its many specialisations; it is narrowly academic, it has much prestige and the most vocal and influential families ensure that schools teach to it. As I have already said, it defines a 'good' school. Alternative paths can thus be so poorly regarded that they barely survive. Rather than permit the growth of alternatives for a larger population, it tends simply to get rid of students. Its diversity is itself or nothing. As a result it must be a prime target for reformers who see increased participation in education as necessary to a democratic society.

This is not to say that reformers are diversificationists. Among radicals there is a minority of do-your-own-thingsers left over from hippie days, but radicals are generally reluctant to promote any form of alternative because they expect alternatives to degenerate into streams. Diversificationists are generally conservatives. What reformers usually want is a Trojan horse — something that will be admitted to the city and will eventually take it over. But there is always a risk that the reform will fail, that it will be admitted to the city but be absorbed by it. The question is whether the situation is so bad that the risk has to be taken.

In the case of the HSC, the Victorian Secondary Teachers Association (VSTA) has thought the risk worth taking now for about fifteen years. Although its campaign against the HSC and against the entry policies of the universities has at least once seriously split its membership, the campaign also appeals to many high school teachers, firstly because their students mostly get the worst deal out of the HSC competition and secondly because many of them have positive impressions of the workings of descriptive assessment in the earlier years of the secondary school.

The VSTA began its assault with a straight bombardment. Tertiary entrance should be open, it declared, with random ballots to decide places in faculties with quotas (Hannan, in press). This caused much excitement but there were no breaches in the wall. In 1975, the VSTA supported a move by a small school — Sydney Road Community School — to set up an

alternative course at Year 12 which would none the less be a valid entry qualification to tertiary institutions (Hannan, in press). Its validity, however, depended on a loophole in the rules governing the HSC, and the controlling body—the Victorian Universities and Schools Examinations Board (VUSEB)—soon closed it. The alternative certificate, known as the STC (Schools Year 12 and Tertiary Entrance Certificate), survived, however. At the time VUSEB was being replaced by a new body with wider powers, the Victorian Institute of Secondary Education (VISE), and people appointed to the new body succeeded in having a two-tier HSC accepted. The old HSC remained and was known as Group 1. The new Group 2 HSC consisted of subjects or whole courses which could be submitted by any course designer to VISE for accreditation and which would be assessed entirely by the school. Many of these Group 2 courses incorporated non-competitive methods of assessment. For many schools it was now possible to have a single form of non-competitive assessment in every year (Victorian Institute of Secondary Education 1978).

So a new dynamic was set up. Inevitably, the Group 1 HSC retained its pre-eminence. Many faculties at Melbourne University and all at Monash University refused to consider Group 2 HSC as an entry qualification. Many schools probably used Group 2 not only as the course for the less intellectually able who insisted on staying, but also as a means of purifying their Group 1 HSC results. None the less, the horse was within the walls.

Styles of assessment

The historical struggles I have described above have thrown up several styles of assessment, each designed to serve particular philosophical purposes and each implying a certain pedagogy. These styles are not easily classified because they often overlap significantly, but, with that reservation in mind, I would suggest classifying them into four:

- graded, or norm-referenced assessment;
- descriptive, or individual-referenced assessment;
- mastery, or criterion-referenced assessment; and
- work-based, or goal-referenced assessment.

I have already described graded a norm-referenced assessment. In its most developed form it places students in rank order on a 'coarse' five- or six-point scale or on a 'fine' percentage scale. The proportions in each grade are fixed according to a statistical model of the distribution of ability. This is the form of assessment commonly found in IQ tests and in achievement and aptitude tests. It is also used to validate tests for literacy and numeracy which yield information about things such as 'reading age'. And, as I have said, it is usual in public examinations, particularly in the HSC. Its aim is fundamentally to compare students one with another, and this information is then often used also to group or select students.

Descriptive or individual-referenced assessment is usually a verbal statement of what the student has done, accompanied by qualitative judgements

from the teacher about the standard of work and the commitment of the student. Comparisons are avoided by restricting summative assessment to an S or N, meaning 'satisfactorily completed' or 'not satisfactorily completed'. Since the descriptive statements normally concentrate on the various elements of the course separately (i.e. on reading comprehension, oral work, participation in classwork, writing, assignments and so forth), the assessment could also be presented in the form of checklists or profiles. The fundamental character of the assessment is that it concentrates on individual achievement, comparing the student's end point with his or her starting point. In this last respect, it overlaps with the next style, mastery assessment.

A mastery or criterion-referenced assessment essentially records the point at which a student has arrived in a set of graded or staged activities. Usually a subject will be broken into a number of relatively discrete activities (as in the profile above), which may or may not be summed up as a global mark. The assessment depends on there being subject-matter which can be coherently staged and tested by behavioural criteria. In some subjects, such as foreign languages, criteria have been elaborated to a point where quite complex behaviour in a simulated linguistic situation may be reasonably assessed. The most common form of mastery assessment, however, is to be found in primary school reading and mathematics kits which aim to exploit the balance between regulation and incentive typical of programmed learning generally. In that context, the itemisation of content and the pace of students are the central ideas of the style.

Work-based or goal-referenced assessment requires that the goals to be achieved are established in advance and that the process by which they can be achieved is stipulated. A subject is divided into units, each representing, say, a term or half-a-term's work. At the beginning of each unit a number of goals are prescribed or negotiated by the teacher and the class. The goals cover both 'participation' demands—attendance, taking part, completing the set workload—and specific 'subject' demands—writing of a certain kind, a range of reading to be commented on, mastery of a set of facts or operations, and so on. If the work is done as specified, that is, if the goals are attained, an assessment of 'satisfactory completion' is given, accompanied by a descriptive statement of the goals. The fundamental idea is that participating students should succeed.

Because of this, for many of its practitioners the essence of this style of assessment is negotiation, which is seen, however, as a group, not an individual process. The style does not necessarily preclude mastery or criterion-referenced assessment as elements in subject goals.

Each of these styles of assessment can be found in practice now in Victoria, from the prep grade to Year 12. Each is a focus of people's hopes for reform or of their fear of it. On the political spectrum, graded assessment is the preserve of conservatives while descriptive assessment is espoused by progressives. Mastery assessment is intended to exorcise the statistical ghost from the grading machine. Work-based assessment is meant to put some intellectual and pedagogic stiffening into descriptive assessment. Of

all four, work-based assessment requires the most radical readjustment of purposes and renovation of subject-matter and pedagogy.

Let us now make a more detailed critique of each of these styles of assessment.

From assessment to evaluation

Once upon a time, the literature on assessment basically addressed itself to the question of what was the most effective form of assessment. The marking of essays was compared with would-be 'objective' testing, the virtues of referencing local results against statewide or nationwide norms were paraded, techniques for comparing marks across subjects were expounded, and the predictive validity of this or that kind of test was discussed.

Today such talk goes on in the wings among the number crunchers. On stage the duel is between protagonists with different purposes, all with their own assessment devices. A critique of these, therefore, depends very much on larger concepts of purposes, equity, intellectual standards and the capacity of pedagogy to effect learning. At the very least, I would want to ask of any assessment style these questions:

- What general social values underpin the style?
- How does the style contribute to the maintenance of intellectual standards?
- Is it an effective tool in classroom management?
- What kind of classroom relationships does it tend to promote?
- To what extent does it lead to a rising standard of pedagogy?

Graded assessment

In the case of graded or norm-referenced assessment I have already implied answers to some of these questions. In its strictly applied forms, norm-referenced assessment takes a deterministic view of human intellectual capacities. Whilst ultimately it elevates individual achievement above all else, its effect through the school system is to consolidate the privileges of certain social groups and to provide some mobility for less privileged groups.

There is a circularity in this process which makes discussion of standards confusing. In a way norm-referenced assessment presumes that those who will do well, will do well. At best, it weeds out the slackers and the genetic throwbacks; otherwise, the 'standards' involved are self-defining—the more elite the top groups, the better the standard. As in a race, only the place-getters or the near record-breakers are excellent. Within its own definition, then, normative assessment maintains standards. The definition can be challenged only by referring the work of the successful to some more absolute standard or alternatively by asking what can be said of the standards of those not in the top group. Proponents of non-competitive assessment would certainly criticise the standards maintained by competitive assessment as being too narrowly academic and too exclusive. But it is hard to arbitrate between the two assessment practices on the grounds of stan-

dards because ultimately the intellectual standards of a course do not depend on the method of assessment in use. The assessment mode is certainly an expression of and an influence on subject-matter and pedagogy, but it is not a determinant of either.

In practice—that is, among teachers and parents (and, often, students)—the great virtue of competitive assessment is said to be its value as an incentive to work. A short while ago I spoke about assessment to a group of Catholic school teachers. At question time, one of the audience put the view that competition was necessary because children are naturally lazy. Another member of the audience jumped in to refute this view as one derived from the concept of original sin, of fallen human nature. Faith in the educability of all children, the second speaker maintained, could serve just as well—indeed better—as an incentive. Philosophically, I would agree with the second speaker. I could also point to substantial episodes in my own teaching experience to confirm it. But the first speaker is also making a valid point, not about morals, which is how it appeared, but about classroom management. Getting a group of thirty kids to work continuously is a rare feat and one of the great accomplishments of teachers. The tradition in that art is largely coercive. One must impose an order, by force of custom and will, and one must police it. ‘Motivation’ is therefore one of the prime preoccupations of teachers, and such is the temper of our times that motivation by force has largely given way to motivation by reward. By now, indeed, some teachers would claim that it has given way to anarchy, that there is not, among some social groups, an acceptance of order strong enough to make any form of motivation effective. If, however, the reward is genuinely valued by the class, competition for it can be an effective form of imposing order. The art would lie in ensuring that the competition was such that nearly all members of the class felt they were within reach of the reward. You cannot have an effective class with a large disaffected group, growing bigger as the pyramid tapers, as is inevitable. The management task then becomes one of either getting rid of them or of distracting them with other rewards. If there are too many to be put in charge of cleaning the yard, the school probably has to develop a more ambitious scheme to remodel and landscape the yard! But that last solution, if it happens to work, opens a Pandora’s box of questions about the nature of academic and practical studies.

In the matter of classroom relationships the individualistic character of competitive assessment tends to dominate. The fact that each student will have an individual assessment—in order for the system to produce a ranking and also an appearance of equity—creates a very strong pressure for classwork to be structured for individual rather than for group work. The taboo on ‘cheating’ becomes very important. Activities such as discussion, where the most competitive talkers may not be the best thinkers, or theatre, which is essentially a group project, or even team sports, become difficult to assess. The most defensible relationship becomes one in which each member of the class is seen as an individual learner making use of the teacher

and of material resources. It is not that co-operative and group activities are discouraged. On the contrary, students are exhorted to value and practise them. They are, however, peripheral to the mode of assessment, and only idiosyncratic students will practise them to any extent.

Since the emphasis of competitive assessment falls so heavily on the capacity of the individual to profit from what the school offers, the mode must be regarded as inherently conservative in its view of subject-matter and pedagogy. This is not to say that schools attached to this mode do not try to improve their curricula. Many do. But the improvement will probably be seen as some kind of modernisation, an improvement for those who already do well rather than a spreading of the course to those who do not. For the latter, improvement will be in terms of alternative courses. Many, however, do not attempt improvement. Far from being attached to the competitive mode, they endure it, fearful that any break from it will penalise those few who do modestly well at it. This is the condition of many schools, both government and non-government, and it produces a lifeless curriculum, lacking in zest and in intellectual enquiry.

In summary, competitive assessment seems to be structurally ambiguous. While it holds out a standard of excellence to an elite, the standard is probably a narrowing and individualistic one. Again, while it is undoubtedly a spur to some, it is probably a deterrent to others. And while it has generally been accompanied by a greater sophistication of academic courses in upper secondary schools, it stands as an obstacle to a thorough renovation of the academic curriculum which would make academic knowledge more useful to a larger number of people. These ambiguities, however, readily make sense when they are matched to some of the most valued instrumental purposes of schooling, especially selection into specialised professional training. Within the confines of that purpose competitive selection and academic specialisation can appear to be necessary even if they are agreed to be unfortunate. One must widen the focus to larger cultural and economic considerations before the ambiguities become contradictions.

Mastery assessment

In the abstract, what I have called assessment for mastery would seem to do the same job as norm-referenced assessment but a lot better. Its social values tend strongly towards the individualistic. Its emphasis on pace through a sequence is meant to favour the high fliers as well as to accommodate the plodders, so in theory it should raise the standard as high as the highest flier can. As a tool for classroom management it is explicitly designed to stop the bright from being bored and the slow from being frustrated. It isolates individuals with the materials and pedagogy to adapt to the full range of individual differences. In principle at least it would seem to be an ideal instrument for bringing all students up to a basic standard in their own time and for subsequently channelling them into a many-tiered job and training market.

In practice, mastery testing has not loosened the grip of grading, in spite

of the fact that versions of it are basic to many elementary reading and mathematics schemes. One reason perhaps is that it is not a good tool for classroom management. Classes in fact stay together and are promoted together so there is a strong will on the part of the teacher, and of parents, to keep them together in at most a handful of sub-groups. There may also be a resistance among teachers to the attempt to itemise and sequence learning, especially when their experience demonstrates that major learning activities such as writing, problem-solving, forming and testing opinions, and experimentation cannot be learnt. I am not saying that assessment of this sort is not used. Especially in primary and junior secondary mathematics and language it is commonplace for teachers to describe and place students by the stage they have reached in some sequenced scheme or other. But when report time comes they generally convert this to a grade or incorporate it into a more global description of the student's work.

Criterion-referencing appears to me to use some of the main assumptions of mastery testing but to apply them to a more sophisticated concept of learning. Its assumption is still that the major elements of learned behaviour can be itemised and observed, but it also assumes that learning is probably best pursued in wholistic situations. In language learning, for instance, it would suppose a pedagogy premised on communication in real or simulated situations and try to provide a set of observable behavioural criteria against which learners can be judged while acting in the situation. Thus it is adaptable to group as well as to individual learning and can be a stimulus to a higher standard of pedagogy. However, if it is to differ from a mere observer's checklist of things to see and grade, it must be based on a very advanced and subtle analysis of the stages a variety of learners may go through in acquiring knowledge and converting it into behaviour. Whether we have such analyses is, in most subjects I know of, doubtful; and if we do, it is certain that they are generally not known in schools or in institutions for training teachers.

Descriptive assessment

Treated purely as a technique, descriptive assessment defies a critique. One must know what is to be described; otherwise descriptive assessments may be simply written out versions of all other kinds of assessments. Historically, however, the intention of descriptive assessment has been to remove competition, and hence to avoid comparisons among students and references to norms. It has therefore been the cutting edge of curricula designed to offer success to all students. Seen in that context, it could be argued that it does attempt to reconcile group interests with individual interest, since although it concentrates on the distance travelled by the individual it is concerned with having the whole group progress and it does not preclude group activity to the extent that graded assessment does.

Gauging the contribution of descriptive assessment to the maintenance of standards also founders on problems of definition. In theory, it is open to any diversity and hence to a raising of standards for some. But it also offers another meaning of raising standards, which is the spreading of success

to a larger group. However, these are theoretical possibilities for which only those who have experience of descriptive assessment over a period of time could answer. My own experience says that standards can be raised under such an assessment method. Those who would do well in any circumstance can do as well or better; those who would *not* do well in a competitive regime can certainly gain more. But these changes are only permitted by the assessment, not brought about by it in any way. The same method of assessment could cover up a drastic slide in the volume of work, in incentives and in intellectual standards—a slide that probably could not occur in an externally moderated normative system. If such a slide is to be prevented, good courses and forms of evaluation which bring teachers together to discuss standards are essential.

Much the same has to be said of the role of descriptive assessment in altering or improving class management, classroom relationships and the standard of pedagogy. So long as the emphasis stays firmly on everyone's achieving some meaningful success, and so long as courses are worked on to that end, descriptive assessment can be an appropriate instrument. Of its nature, however, it merely opens the space—it does not suggest how it might be filled. The focus of our critique, in other words, has shifted from ways of judging the student to ways of judging the course.

Work-based assessment

Work- or goal-based assessment makes this last distinction sharper. Since it is historically an attempt to tighten up the principles underlying descriptive assessment, it has a similar view of the need to balance individual and group purposes and to raise standards by extending horizontally as well as vertically. It is, however, more systematic in providing incentives to ease classroom management, in establishing participatory relationships and in insisting that the pedagogy must ultimately carry the responsibility of achieving a worthwhile measure of success for students.

The incentive used in goal-based assessment is basically the old pass/fail one. You earn an assessment of satisfactory completion only if you do the work as stipulated in the goals. The fact that you may also have taken part in formulating the goals can be an added incentive to engage in the work; but, either way, you have to do it in order to pass. You may be brilliant, but if you do not complete the course, that is what your assessment will say. In some cases, the ability paradigm that is integral to norm-referenced assessment, will be turned on its head. An A for achievement and E for effort no longer apply. Basically, what you need is S (satisfactory) for effort.

Such a proposal immediately draws two responses: its basic fairness is conceded (the students are no longer the culprits), but what if their work really is bad? In some cases, of course, that question will be unanswerable, and a particular judgement will have to be made taking into account the use to which the assessment is to be put. Such cases do, in fact, crop up in all systems—people fail but are selected anyway, bad marks in one part are discounted against good marks in another. Generally, however, the answer has to lie in the course and in the pedagogy. If it is fair to pass

students who have conscientiously participated, it is necessary to ensure that what they have participated in will work. The goals have to be both educational and attainable. The work to be done has to be both worthwhile and doable.

In practical terms this means that the goals have to specify not only the outcome — which is what criterion-referencing does — but also the means by which the outcome is attained. The distinction is akin to that between aims and processes. The aim is a statement of where you want to reach — to be able to write a story, for example. The process, or goal, is a statement of how you are going to go about reaching that aim — by talking or thinking it out, writing a draft, going over it with someone else, correcting it, editing it, redrafting, and so on. Similarly, if being able to take part in making decisions responsibly is an aim, then taking an active part in making certain decisions during the unit is the goal. Do that in the agreed manner and you pass. If you are dissatisfied with the outcome, you will look not only to weaknesses in the student but also to weaknesses in the aim or in the process. Thus there is an inbuilt pressure, especially when the course is reviewed by a group of teachers, to reconsider the pedagogy of the course when outcomes do not appear to be fulfilling the aims. There has been a shift from assessment of students to evaluation of courses.

I will finish by looking at some of the conditions for effective evaluation, but before that we should note that descriptive and goal-based modes of assessment also introduce a new element to assessment, namely student self-assessment. In the goal-based mode, self-assessment starts when the goals of the course are set in the context of discussion by the whole group. This negotiation process applies especially to 'participation' goals (how many absences are permissible, for example) and may extend to choosing among or proposing 'subject' goals. Obviously the possibilities depend on the nature of the subject, the maturity of the students and the purpose of the course. If then the process is made clear and the aim understood, the final assessment can also be open to negotiation, with student and teacher comparing their assessments as a basis for reaching agreement, and referring to another person or group for arbitration, if necessary. If such a system can be made to work, classroom relationships are fundamentally altered and a fertile basis for course evaluation is established.

At this point we could consider briefly the connections between various styles of assessment and evaluation and the overall structures of institutions. I have already noted the links that exist historically between the practice of grading and the practices of streaming and tracking. I have noted, too, that the differentiations implicit in these practices have been expressed in institutional structures at the systemic level, at least in secondary schooling. The hierarchy of district comprehensives, selective high schools and elite private colleges could be seen as parallels to streams within a school; whilst the distinction between academic schools and technical schools could be seen as the systemic parallel to tracking.

During the 1960s more sophisticated variants of streaming appeared in both primary and secondary schools. The Wyndham Scheme in New South

Wales, for example, encouraged setting and banding, which took students out of more or less permanent ability streams and placed them in various groupings according to the level they had reached in each subject. In the primary school, graded materials appeared in kits (the classic example being the reading kits of the SRA) which promised that not merely groups, but individuals could be organised to work at their own pace. It is probable that both of these developments were actually intended to reduce the undesirable effects of more rigid forms of streaming, in particular the 'self-fulfilling' failure attendant upon being placed permanently in a bottom stream. Basically, the new techniques implied either that recovery and remediation were possible or that students would at least obtain more satisfaction from the experience of steady success that working through graded material at one's own pace promised.

During the 1970s the idea of working at one's own level was married to ideas about students' interests and the relevance of content. The notion of core plus options, which had always existed to some degree at some levels of schooling, was applied more widely, so that in the so-called basics students would find themselves placed at a level said to be appropriate for them, whilst in the non-basics (social studies, arts and so on) a range of topics or activities would be offered from which the students would choose. Thus student choice came to be a factor in deciding differentiation.

This notion of choice, based initially on a fairly simple idea that students would be more satisfied and more successful doing what they, in some sense, wanted to do, was destined to have large structural repercussions. Traditionally, state schools have been district schools, the exception being the handful of academically selective high schools, such as Melbourne High School and Fort Street. The middle of the 1970s, however, saw this idea challenged. The Commonwealth Schools Commission set up a project on choice and diversity in government schooling, and the Liberal governments of the time borrowed the idea of choice also to rationalise the continued existence, in fact the growth, of private schools. Choice, in the market sense of the term, became a strong element in deciding systemic structures.

Once governments began to publicise the idea of choice, as happened in South Australia for example, pressures immediately developed for schools to try to differentiate themselves, and of course, for parents to seek out what they understood to be the 'best' schools. The result was a powerful reinforcing of the status of those schools which promised the obviously popular things: good order and good results. Since results come mainly in the form of marks for academic work, this meant in effect that schools claiming to be academic rose to the top of a hierarchy above the district comprehensive and well above the struggling school in a low socio-economic area trying out a few alternatives.

All of this may seem a long way from assessment questions. In a way it is, because we are discussing not so much a chain of causes and effects as a set of practices that fall together historically and share common ideologies. In everyday discussion about schools, however, assessment is strongly linked to these larger institutional rearrangements. A school which

sees itself as having to tighten up its image in order to hold its numbers does promise that it will give regular gradings, and that it will stream so that the best of its students will be able to move ahead at the maximum pace. It may promise regular testing. It will introduce the courses it believes get the best results in whatever examinations system happens to prevail. In short, it will announce a policy of assessment (grading), and one of evaluation (courses and methods will be regulated by publicly tested results). This will put it in significant contrast to a school claiming widespread success for all its students irrespective of their ability; promising that it will adjust its courses to its students; deliberately taking all comers; and working on the evaluative principle that courses have to be adjusted to correspond to the present needs of all students. Inside each of these schools we will see fundamentally different possibilities. In the former, the idea that students must measure up to fixed standards within pre-defined strata is paramount. The basic relationships within the institution will tend towards coercion, so that students will be made to fit the structure. In the second, we will find a more uncertain pattern of relationships resulting from the attempt to adapt structures to students. In the terms of this monograph, there is a shift of focus from assessment in the first to evaluation in the second.

Evaluating courses

If a shift of emphasis from assessment of students to evaluation of courses is to succeed, some new conditions must be created. There is little to be gained from grafting one form of assessment on to a course designed for another. Course designs and assessment modes do match and reinforce one another, especially, of course, at the level of pedagogy.

The first and most obvious of these conditions is that schools be organised so that teachers of the same or similar courses can meet. To me this seems to be the essential pre-condition for course evaluation. Certainly, it is all to the good if course designers and parents and students can also have some part in evaluation, but the core of the process is the pooling of practitioners' ideas. A good basic course design does, of course, make the evaluation process more efficient and the course itself more stable and hence more likely to inspire confidence in its users; but it cannot be emphasised too strongly that a very large part of course evaluation is evaluation of pedagogy, and it is in pedagogy that teachers excel. A good course design, therefore, will provide for teachers to consult one another and compare methods, just as a good school will.

The next condition is that the course under evaluation be clearly placed in its context. Schooling is a strongly staged process — perhaps too strongly — during which people expect to develop in knowledge and skill. Students move in age grades and pass through several transitional points. Every two or three years the question of what stage students have reached has to be faced. No teacher can safely be blithe about students working away at what they can or will. Every so often, someone asks: 'Where are they up to?'

Interestingly enough, every style of assessment has ways of fudging the answer to this. Students can be given good grades 'for the course they are taking' (which happens to be terminal). A pass (D) can turn out to be an unpromotable mark. A descriptive assessment can give students good reports without detailing what it is they can do. A goal-based course can be 'satisfactorily completed' but ineffectual for learning. The fact that most styles have their way of fudging the issue shows that the problem of progress is widespread. But to set it aside with relativistic statements about students doing their best puts us on a very slippery slope. Certainly students should do their best. But courses have to fit that best into a coherent context. Students are entitled to know, in such an important matter, where their best puts them. Although it is an awkward problem for course evaluators using any kind of assessment, it is a necessary constraint for any evaluation to work within.

Beyond these two conditions, of regular discussion among teachers within a realistic context of constraints, there is another which really engulfs all considerations. It is that evaluation be premised on the possibility of success. Basically this means that evaluation starts from the assertion that subject-matter and pedagogy can be conceived in such a way that the course is accessible to all qualified and participating students. And by 'qualified' here I mean not that they have passed attainment tests for entry, but simply that they have completed whatever courses lead up to this one in the normal progress through schooling (the 'coherent context' condition again). Obviously, in particular schools or classes, this can be a difficult assertion to stick with because for many reasons many students will not be qualified and will not be able to or want to participate. If their numbers are high, the whole school's program will be at risk. But even in those desperate circumstances, there is a case for holding to a stable coherent practice as the norm to which special programs can be clearly related.

This overarching condition needs explaining. Some form of evaluation is possible in any system. What will vary is the extent to which it can be carried out by the participants and in the interests of all students. Externally prescribed systems, for example, permit little more than generalised agitation from practitioners. Competitive systems have trouble catering for all. My assertion here is that non-competitive systems can permit more productive evaluation than can competitive ones; and that systems premised on the educability of all students will be the most productive. We must, however, attend to the meaning of educability, and in particular relate it to the question of commonality in curriculum.

This brings us to a definition of the notion of success, which is the critical notion at stake in debate about assessment and evaluation. Competitive systems necessarily deny that significantly equal outcomes can result from engagement in common activities. No competitive course can offer a plausible measure of success to all its students. Either there will be a group of outright failures, or there will be a sort of *lumpen* group for whom the course or the perceived standard will have been watered down. The directions for improvement, therefore, tend to be towards increasing the

sophistication of the existing course for the successful and multiplying the range of alternative courses and tracks for the unsuccessful. Further, the degree to which these various tracks or streams need to be evaluated at all for their appropriateness and their success is limited by the fact that students may be excluded either directly or by being guided into a transfer. Commonality, in short, is not an element in the definition of success.

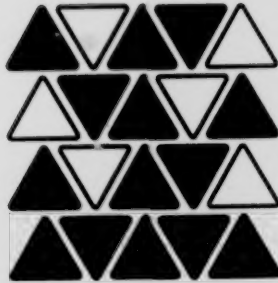
A non-competitive system will not of itself overthrow that kind of thinking. Differentiation of courses is still a basic concept within many descriptive and 'distance travelled' systems of assessment. However, non-competitive systems do make it possible to premiss courses on commonality rather than on differentiation, and hence on a conception of success that is not relativistic.

The reasons for wanting to adopt such a notion of success lie outside the scope of a discussion of assessment. They are reasons that must be derived from cultural arguments. We have set up school systems for both immediate instrumental and longer term cultural reasons. We mean schooling to contribute usefully to people's lives as individuals, as social beings, as workers and as citizens. Most of those purposes suggest a high level of common cultural formation, rather than an individualistic and stratified formation, though none is entirely unambiguous. Success in schooling has to be judged within that context. To the extent that we choose to have a society in which all people have common cultural concerns and some significant power to participate in social affairs, we will see a need to change from competitive modes to more inclusive ones both in the assessment of students and in the evaluation of courses.

References

- Dow, G. M. (1970), 'Streaming', *Secondary Teacher* (157), July, 21-8.
 Grant, J. L., & Hannan, W. (1981), 'Principle vs. principal', *Secondary Teacher* (5), *TTUV News* 15 (4) (joint issue), 14-24.
 Hannan, B. (1982), 'Assessment: A workable democratic method', *English in Australia* (59), March, 11-19.
 Hannan, B. (in press), *The Democratic Curriculum*, George Allen & Unwin, Sydney.
 Houts, P. (ed.) (1977), *The Myth of Measurability*, Hart, New York.
Secondary Teacher (1970) (157), July.
 Victorian Institute of Secondary Education (1978), *Policy Statement: Curriculum and Assessment in Year 12*, VISE, Melbourne.

Readings



1

Educational versus managerial evaluation in schools

Richard J. Bates

Of the three message systems of schools — curriculum, pedagogy and evaluation — the evaluation system is the most neglected and the most in need of evaluation. Even Bernstein's (1975) analysis concentrates almost exclusively on concepts of classification and framing — concepts associated with curriculum and pedagogy respectively. This lack of attention is partly due to the identification of evaluation with the testing movement and the resultant mystique arising from statistical complexities. Such intricate treatment of day-to-day problems simply baffles many teachers. It also seems that the ritualization of evaluation systems in schools is so commonplace, so constant and so pervasive that evaluation becomes taken for granted by teachers and pupils alike. The lack of challenge to evaluation practices often results from the emphasis of most school evaluation systems on the evaluation of pupils. That is, evaluation is typically conducted by the powerful on the powerless. Another part of the power of evaluation systems derives from the difficulty of imagining alternatives, for, as Durkheim suggests:

We are so accustomed to believing that emulation is the essential motive force in academic life, that we cannot easily imagine how a school could exist which did not have a carefully worked out system of graduated rewards in order to keep the enthusiasm of the pupils perpetually alive. Good marks, solemn statements of satisfactory performance, distinctions, competitive essays, prize-givings: all these seem to us, in differing degrees, the necessary accompaniments to any social education system. (Durkheim, 1930, trans. Peter Collins, 1977, p. 159)

The tangible system of inducements and rewards by which students are thought to be motivated is a visible message system through which effort and achievement are recognized and rewarded. Such is the conventional wisdom of schools. Within it several ideas are brought together. Three ideas seem to be most important: diagnosis, motivation and reward. The fundamental formula of the meritocracy is, after all, IQ plus effort = merit. Diagnosis of ability

according to IQ or some other appropriate measure, assessment of motivation which defines effort and a carefully structured hierarchy of rewards, acknowledge publicly the honour due to the meritorious. On the basis of such fine distinctions students can subsequently be recognized by the wider society and appropriately located in the hierarchy of opportunity, status and reward.

What is immediately noticeable about this description is how the supposed educational purposes of the evaluation systems employed by schools are rapidly subsumed by the wider processes of social allocation and the creation of hierarchies. Rather than a means of furthering the *educational* purposes of individual development and emancipation, educational evaluation has become, like class, a mechanism of exclusion (Bernstein, 1975). The power of the evaluation systems employed by schools to achieve this exclusion is one of their most noteworthy characteristics (Broadfoot, 1979, 1980). Indeed, at a number of crucial points within the educational careers of pupils, school evaluations are employed to deny further educational opportunity to certain categories of pupils. Thus restrictions are placed on access to the world of meaning. As Musgrave argues, the prevalence of such mechanisms seems to indicate that 'many of those with power over the education system would seem to be afraid of the results for them if the masses came to understand more meaningfully' (1979, p. 100).

School evaluation systems are not, however, self-contained, self-sustaining systems linked to the outside world only via the careers of their pupils. The development and application of standardized tests on a national and international basis allows wider evaluations of educational performance. It also allows, indeed insists upon, an individualistic ethos which

... provides the rationale for prevailing modes of educational discourse, and whether the subject is a pupil's progress, school and classroom organisation, unit costs or teacher inspection, evaluation — accountability on an individual basis — permeates every aspect of systemic functioning from provision to practice (Broadfoot, 1980, p. 6).

The effect of the individualistic ethos is to make individuals accountable to their institutions rather than institutions accountable to their members. Indeed, the introduction of evaluation systems of this kind was an accompaniment to, if not an instrument of, the removal of local accountability between schools and their communities and its institution into the governing bureaucracies of mass systems. This point is made by Broadfoot who claims that '... the institution of formal evaluation procedures within education was directly instrumental in rationalising educational provision into a system' (1980, p. 6; see also Katz, 1975; Tyack, 1974).

The evaluation systems employed by schools seem, therefore, to serve several purposes.

- 1 They are an important part of curricular and pedagogical processes in

terms of feeding back information to pupils, teachers and parents on the efficacy of school practice and pupil learning.

- 2 They also serve the wider purpose of social allocation, stratification and exclusion.
- 3 The ideology they convey is one of individual rather than institutional accountability which incorporates judgements of worth and blame into the consciousness of teachers, pupils and parents alike.
- 4 They allow the development of mechanisms of hierarchical accountability and control and enhance the steering capacity of centralized government.

There are obvious tensions between these various demands. For instance, the educational benefits of evaluation of curricular and pedagogical processes may be at variance with the school's need to 'project a positive public image'. Similarly, the school's agency in social selection and allocation may run counter to pupils', parents' and communities' aspirations. Again, the accountability of institutions to their clients may well be disguised by the employment of evaluation practices directed towards competitive individual assessment.

Such tensions as result from these various demands involve the differing interests of groups such as pupils, parents, teachers, administrators, planners, economists and government. This paper outlines the way in which various evaluation practices embrace or deny such interests. In particular, it argues that two distinct but contradictory aspects of school evaluation systems can be identified: those relating to managerial and to educational interests and purposes.

Classification versus Dialogue in the Evaluation of Pupils

School evaluation systems which assess pupil achievement in order to establish and contribute to a dialogue over teaching and learning clearly need to be built on a consideration of the various options available in the use of tests. Some of the key distinctions between these options are set out by Stodolsky (1975), who argues that while traditional psychometrics have emphasized summative, norm-referenced, performance and standardized tests as the appropriately 'scientific' and 'objective' measures of pupil performance, and while such tests (despite their low levels of predictive validity) serve purposes of selection, classification and allocation of pupils, they are much less useful in guiding the instructional and educational processes that should be the focus of classroom activity.

By the very nature of the way they are constructed, however, these tests *cannot* provide the child, his teachers or parents information about what he has *specifically* learned or not learned in a given subject matter over a period of time. Since standardised tests are only samplings of course material, scores cannot be used to determine

explicit instructional needs of children in any but the most global way.
(Stodolsky, 1975, p. 14)

What is needed, suggests Stodolsky, are tests that are more clearly and usefully linked to educational purposes and environments and more capable of guiding pupils, parents and teachers. Here, she suggests, the need is for formative, criterion-referenced, diagnostic and growth-centred testing procedures which 'are meant primarily as aids in the instructional process and can be related much more closely to the classroom learning of the child' (Stodolsky, 1975, p. 16).

Most school evaluation systems are, however, premised on a certain scepticism concerning the effects of teaching on learning. Good teaching, elegant teaching even, may be a pleasure to watch but, ask the sceptics, does it bring about learning? The obvious way of answering critics is to ask the pupils. Whether directly or indirectly gathered, the only information appropriate to a judgement of the effectiveness of what schools do is information derived from pupils. Asking pupils for feedback on their learning is not, however, a straightforward issue. Even a simple set of teacher or observer constructed questions concerning, for instance, a particular lesson may contain unsuspected difficulties. For instance, it may include questions which

- 1 relate to ideas or knowledge that some children already possessed before the teaching took place (so is there no learning?);
- 2 use language that is too complex for some children to understand (even though they may have learned what the teacher intended);
- 3 focus on areas of the lesson where no learning took place (although other kinds of learning might well have occurred).

The literature of testing is strewn with complex, ingenious and only partly successful attempts to answer such problems. Indeed, the complexity of the technology that has arisen in this area often obscures, if it does not replace, the *educational* purposes of asking such questions in the first place.

The educational reasons for asking such questions are fairly simple. Teachers need to know what their pupils already know (so that they can build upon it). They need to know whether their pupils are learning what they are attempting to teach (so that they can modify their teaching accordingly). They need to know what else their pupils are learning (so that they can draw the unexpected into their consciousness).

Normal testing procedures do not provide such information in a sufficiently rich and comprehensive form for it to be of substantial assistance in the improvement of teaching. The major problem with such procedures is their fragmentation of knowledge, experience and understanding. For evaluation processes to be of real assistance to teachers a more holistic approach is required — one which is securely located in the actual classroom experience of pupil and teacher. This is especially necessary in informal educational environ-

ments where the standardized instruction best suited to standardized tests is absent and more complex forms of teaching and learning are employed.

A variety of possibilities exists for alternative forms of evaluation. The systematic use of work samples, observational schedules, and interviews with children are all promising methods (see Hawes, 1974; Codd, 1981). What is characteristic of the information derived from such procedures is that it not only assists the teacher to enrich and develop his own teaching but can also become the basis for an educational dialogue with pupils and parents concerning what is being learned and how best to provide various kinds of learning in the future. This kind of dialogue in individual cases inevitably spills over to the consideration and evaluation of the wider contexts of learning provided within classroom and school. Such considerations necessarily involve collaborative explorations of the context, criteria, comparisons and judgements to be made with regard to the learning environment (Bates, 1981). The resultant dialogue is itself a process of continuous evaluation.

Frequently, however, the place of such dialogue is taken by forms of evaluation which are technical, rather than illuminative. In the search for objectivity and for 'accurate' information regarding pupil learning, formal testing procedures are substituted for dialogue and discussion. As a supplement to dialogue, testing of certain kinds may be helpful. Tests may well be a quick and rough guide to what is going on in the classroom. It is important, however, that the tests used are appropriate for the learning that is supposed to be occurring. For instance, if, as Berliner (1980) recently suggested, the two most commonly used standardized maths tests in the USA test only 20 to 30 per cent of the curricular material included in the four most commonly used maths texts, then they are clearly of limited use in assessing the outcomes of instruction based on these texts. Nevertheless, testing programs based upon such silliness are not only used to make political judgements about standards but also to make decisions concerning the educational futures of individual children. This is a far cry from the initial intention of the teacher.

It is clear that tests can be used in a number of different ways. On the whole, formal tests may well serve *educational* purposes (that is, provide information about the *processes* of learning in the classroom) if they are formative, criterion-referenced, diagnostic and teacher-made. Such tests can support, develop and extend the teacher's awareness of what is happening during dialogue with pupils and deepen the complexity, richness and usefulness of that dialogue for teachers and pupils alike.

On the other hand, summative, norm-referenced, performance and standardized tests tend to serve other, *managerial* purposes (that is, purposes having to do with processes of sorting, classification, allocation and control). Such tests do little to further the processes of learning in schools and classrooms. They have far more to do with maintaining the *status quo* in the wider social structure, an issue that we shall return to later.

For the moment, however, it is sufficient to point out the tension that exists between these two kinds of evaluation and the rather fundamental

decision schools need to make with regard to the particular kind of evaluation system they implement.

The evaluation system adopted by most schools tends towards the managerial model. In such cases the techniques of evaluation employed can have rather unfortunate, anti-educational, consequences. For instance, while much of the discussion of evaluation in schools refers almost exclusively to the evaluation of cognitive gains, school evaluation is not focused simply on cognitive processes and achievements. Indeed, as Jackson (1968) makes quite clear, evaluation in the classroom context is a continuous, public and all-embracing process covering not only cognitive but also social, affective and moral behaviour. Such non-cognitive evaluations are an important part of school evaluation systems having, as Parsons (1959) suggested, a major influence on the socializing process. Proper and improper behaviours are often clearly defined by the school through its rituals, rules of participation and exclusion and its labelling processes. While Rosenthal and Jacobsen's (1968) study of the 'pygmalion effect' was centred on cognitive performance, there are good grounds for believing that the effect is even more powerful in the non-cognitive areas (Waller, 1967; Bernstein, 1973; Sharp and Green, 1975).

The effects of labelling on the development, even promotion, of officially 'delinquent' behaviour have been documented in a variety of schools (Hargreaves, 1967, 1976; Lacey, 1970). Quite often the labels follow the grouping of pupils on the basis of certain evaluations of their ability and potential. Equally often such groupings are closely related to other social and behavioural factors. The result is that the school evaluation system assists not in the promotion of *educational* ends but in the achievement of certain more or less desirable *social* ends: the creation and support of sub-cultural structures within the school which are closely related to, and serve to reproduce external social and cultural distinctions. Willis's (1977) discussion of the educational careers of 'the lads' is a case in point. In this case, despite the lads' partial penetration of the social and economic structures and despite the school's explicit promise of possible escape from disadvantage, the cultural forms of teaching and learning employed and the power of the cultural myths reinforced by the school act to perpetuate their disadvantage. In such ways school evaluation systems often support, not educational aims, but social and political processes that turn the school away from its avowed aims of development, opportunity and liberation and towards a role as a 'repeater' of existing forms of social relations.

The Managerial Evaluation of Pupils

Broadfoot claims that assessment and evaluation in schools should be seen 'as an instrument of control, helping to reproduce the inequalities in society' (1979, p. 117). This somewhat overstates the case as not all assessment and evaluation is directed towards that end. Indeed, the previous discussion attempts to justify a distinction between *educational* evaluation, which is

supportive of dialogue and development in the classroom, and *managerial* evaluation, which is directed towards classification, discrimination and control. If this distinction is accepted, then educational evaluation can be exempted from intimate association with processes of social control. It may, in fact, serve some of the purposes it claims, of individual development, of social and cultural enrichment, of liberating the imagination and of promoting concern for truth, justice and beauty. Managerial evaluation, on the other hand, may serve to restrict such possibilities in the name of social order. In particular, it may well be seen as an instrument for increasing control over both the sorting and allocation of individuals into the socio-economic structure and over the distribution of schooling.

Managerial evaluation is very much concerned with the sorting and allocation of individuals into the bureaucracies of government, business and industry. This is an explicit function widely recognized in most Western systems of government. For instance, the Williams Report (1979) placed considerable emphasis on this function of schools. It is equally true of American and British systems where

A great deal of energy is devoted in the school system to 'sorting' children, though less at the age of eleven than in the past, when in this country (U.K.) a universal and traumatic sorting used to take place. The public examination system . . . is essentially a system of sorting, in particular for the filling of places in higher education which still do not fully match demand and for the selection of candidates for posts in business and industry, where competition is frequently fierce. (Kay, 1979, p. 19)

During periods of economic expansion the sorting process could claim to be a mechanism by which the existence of talent within the society could be roughly matched with the expanding occupational structure. Even during such periods, however, there is little doubt that many individuals were 'overqualified' for the jobs they held. For instance, even in 1961 when complaints were being made that Britain was grossly short of science graduates the census showed that 'out of a total of about 260,000 scientists and technologists, 50,000 were in jobs that did not make full and direct use of their qualifications' (West in Spring, 1980, p. 70). The sorting mechanism is obviously very approximate, even in good times. In bad times the mismatch between achievement and opportunity becomes greater.

A major educational crisis of the late 1970s was the increasingly high educational level of the unemployed. The better educated began to encounter economies with slow rates of growth and expansion. The supply of educated people began to exceed demand (Spring, 1980, p. 55). The response to this crisis was that the unqualified sought qualifications while the qualified sought higher and yet higher qualifications. 'The paradox of the situation is that the worse the educated unemployment situation gets and the more useless

educational certificates become, the stronger grows the pressure for an expansion of educational facilities' (Dore, 1976, pp. 4-5).

This pressure for expansion fed the second managerial problem — that of increasing costs. As early as 1968, Coombs had predicted the educational crisis of the 1970s, analyzing it in terms of an increasing social demand for education confronting increasing educational costs and unemployment for school graduates (Coombs, 1968). The prediction was correct although the crisis has been made even more severe by the reorganization of industry through the multinational manipulation of the 'world economy' and the introduction of new technologies in fields such as communications and machine-control technologies (cybernetics and the micro-chip for instance). The overall impact has been an escalation of educational graduation rates and costs at all levels combined with a decrease in employment opportunities in virtually all developed countries of the Western world.

The response of government managers has been to attempt to dampen the demand for education, especially higher education, by withdrawing state funds, by transferring costs to parents or, most significantly from our point of view, escalating admission requirements (see Carnoy, 1977). The escalation of admission requirements to tertiary education has a washback effect on all other levels of education. One of the immediate effects is to produce a general belief that standards at lower levels are falling as a result of assertions that, for instance, 'fewer students met the requirements for university entry this year than last'.

The call for higher standards and a return to basics follows and schools are put under considerable pressure to display their competence and the adequacy of their pupils' levels of performance. Frequently, attempts to do so on the part of individual schools result in a move towards testing procedures — especially testing procedures which are assumed to be scientific, objective, impartial and independent of individual teachers or schools. The result has been a bonanza for test publishers.

Thus, for example, each year some 95 per cent of American high school children take an ability test. As a result of this in 1975 one hundred and fifty million dollars were spent in the USA on standardized test programs. The bulk of this market is cornered by the Educational Testing Service with a turnover of fifty three million dollars in 1975. Harcourt Brace Publishers follow with a turnover of some twenty million dollars (Galton, 1980, p. 137). Testing is big business and is frequently linked with universities or foundations which provide the expertise and the prestige necessary for successful sales. For instance:

One of the outstanding test developers, E.F. Lindquist, has to his credit the two most popular test batteries, the Iowa Test of Basic Skills (ITBS) and the Iowa Test of Educational Development (ITED). Each test is placed with a different publisher which helps to keep the sales of the two competitive. A grateful industry has built a three and a half

million dollar 'Centre for Measurement' in his name at the State University of Iowa (Galton, 1980, p. 137).

The situation in Britain, while not so extreme, is certainly similar. The National Foundation for Educational Research is, for instance, vigorously marketing a variety of tests to local educational authorities wishing to evaluate their schools and defend their standards. In Australia similar trends are apparent with 70 per cent of the income of the Australian Council for Educational Research being generated from test sales during 1979-80. As Hein (1975) suggests, tests and their use are deeply embedded in the fabric of society.

The use of standardized tests on such a widespread basis has little to do with their possible formative, diagnostic or developmental possibilities even though many of the tests 'were originally designed for diagnostic purposes rather than for standardised comparisons' (Galton, 1980, pp. 137-8). The manner in which such tests are applied also militates against their use as a means of educational evaluation. For instance, most of the American tests are scored, not by teachers, but by testing centres often linked to universities or statewide agencies. As Galton observes:

Total secrecy surrounds the construction and computer marking of such tests. . . . Currently the development of electronic scanners to score as well as mark the tests by computer has created added interest within the computer industry itself and more and more royalties are now pouring into the universities to support this work (1980, p. 137).

As the construction, marking and scoring of such tests becomes remote from the control of either test constructors or teachers, the potential of such testing for *educational* evaluation declines. Simultaneously, the *managerial* potential increases as the 'objective' nature of the now reified data allows 'dispassionate' and 'scientific' comparisons between pupils and schools. These comparisons have their political benefits as they apparently allow the manipulation of system inputs, the exertion of sanctions over the internal processes of schools and the ordering of system outputs. As Hein suggests:

The major focus of the present testing programs is not to determine how much children know, to diagnose their learning stages, and to assist them in their growth and development but, rather, to sort and classify children for their assigned roles in society. For this task the present system, with its shoddy and discriminatory tests works quite well and almost independently of the quality of the tests themselves (1975, p. 29).

The school's classification of individuals for managerial purposes on the basis of such tests does not require such elaborate rituals although they do serve to obscure and mystify the political process involved. Similar rituals

borrowed from the scientific community have been available to educators for a long time. One such ritual involves appeals to the scientific use of the 'normal curve of distribution' which suggest that random distributions of virtually anything tend to conform to a bell shaped curve which can be dissected mathematically into standard deviations from the means and used to plot the location of individual items with respect to the mean. The model and its assumptions have been used frequently for the classification of educational performance. The most common version is one which divides the normal distribution into five areas corresponding roughly to the statistically predicted standard deviations and labels them A-E or 1-5. These labels are then used to grade children's performance, such a grading being justified by appeal to the scientific form of the random distribution curve. Subsequently, the precision of the distribution often degenerates into broad categories labelled A — outstanding, B — very good, C — good, D — satisfactory and E — not satisfactory. These labels obviously allow a great deal of latitude in interpretation — especially, for instance, category D which might well mean 'below average' in statistical terms.

There are objections to the use of this procedure even in its more precise form. Not the least of these is that education systems are designed to *interfere* with the random distribution of ability and performance that might be assumed to occur in an untutored population. These arguments are, however, somewhat irrelevant for our purposes. The important point is that such a classification scheme exists and can be appropriated and legitimized by appeals to professional and technical expertise. It can, therefore, be used to classify, select and allocate. Its accuracy is, like the 'shoddy and discriminatory' tests that Hein refers to, largely irrelevant. The point is that it can be, and often is, used and legitimated in the cause of managerial evaluation.

The conflict between such managerial uses of evaluation and the processes of educational evaluation is obviously a tense one. Most of the time it can be 'managed' through the processes of mystification and the exercise of authority. Occasionally, however, the logic of such managerial uses of evaluation is challenged. Most frequently this is done by parents in regard to the justice and accuracy of external public examination results. Less frequently public challenges are made by teachers on grounds which pit the educational uses of evaluation against managerial uses. Such instances clearly reveal the association of managerial evaluation with processes of social, institutional and bureaucratic control.

A Case Study: Managerial versus Educational Evaluation

An interesting case of such conflict between educational and managerial evaluation occurred in a high school in Melbourne, Australia in 1980 (Grant and Hannan, 1981). The conflict pitted a majority of the school's teachers, supported by their union, against the authority of the principal, supported by

the hierarchy of the State Department. The issue was represented by the teachers as one in which their professional independence (in terms of their right to collective self-determination of certain aspects of school policy) was challenged by autocratic attempts by the principal to impose an educationally unjustifiable requirement. The requirement was that staff pursue a policy of regular competitive assessment and report the results as grades from A-E representing a range of percentage marks.

The staff claimed that school policy on assessment had been determined at a staff conference called by the principal in July 1978 at which there had been a decision, carried by fifty votes to seven to replace the presently existing competitive assessment policy with one of descriptive, non-graded assessment. The principal voted against the decision and subsequently in April 1979 issued staff with notice of school assessment policy which ignored the result of the staff conference. Staff and faculty meetings reiterated their support for the decision of the staff conference to use non-competitive, non-graded assessment procedures. On 6 June the principal required each teacher to inform him 'personally of his/her intention in regard to this matter by Friday 8th June'. The Staff Association met and requested that the principal reopen the discussion in a manner consistent with the previous July's staff conference. The principal refused and countered with a note in which he stated: 'I now find it necessary to restate the function of the Staff Association: to deal with staff social matters.' Following this rebuff the matter was taken up by staff, meeting as a branch of their union, with the result that half the school staff declared that, in keeping with the decision of the staff conference, the policy of their union and the declaration of the State Curriculum Advisory Board that 'there is no place for competitive assessment in secondary schools', they would adopt a policy of non-competitive assessment. The principal subsequently reported twenty-two of his staff to the Director of Secondary Education for not completing student reports 'according to school policy'. Inspectors were called in. They interviewed two teachers and instructed them to convert their descriptive reports to a letter grade system. The teachers' union discussed the matter with the Director General without conclusion. In 1980 teachers were twice more ordered by the principal to convert their detailed comprehensive written reports on students into letter grades on mark sheets. Eighteen teachers refused, claiming that giving grades would conflict with the purpose and rationale behind their descriptive, non-graded assessments. On 1 December these teachers received individual letters from the Director General of Education charging them with a breach of the Teaching Service Act 1958 in that they 'did fail to obey promptly the lawful instruction which was given, "namely", to enter on the Form Mark Sheets an assessment for each pupil ... using the rating scheme: A — Outstanding, B — Very Good, C — Good, D — Satisfactory, E — Not Satisfactory.' Following some administrative confusion the Minister referred the matter to the Teachers' Tribunal which has power to reprimand, fine, deprive of increments or dismiss teachers. In February 1981 the Teachers' Tribunal began to hear the charges.

What was immediately obvious from the unofficial transcript of the hearing (Grant and Hannan, 1981) is the clear conflict of principle between the teachers' advocacy of the value of full descriptive evaluation and the principal's insistence on competitive graded assessment. In our terms, the conflict was between principles of *educational* as against *managerial* evaluation. The principal's insistence on the adoption of a managerial form of evaluation was based essentially on two grounds: (1) that it was bureaucratically convenient: it is the 'quickest way' of responding to requests for information from welfare agencies and potential employers, of deciding whether a student has educational problems, and (2) that as a principal, he had the authority to require any teacher under his supervision to obey promptly any instruction until such time as such an instruction should be amended or queried by the Director General as a result of an appeal. The teachers argued that a comprehensive system of descriptive assessment was more educationally defensible and meaningful than a highly abbreviated form of competitive assessment represented as letter grades, the meaning of which was ambiguous.

At the end of the hearing, the barristers summed up these arguments as follows:

Mr. Impey for the teachers argued that the Principal's instruction could have been complied with only in a base physical form. Anyone could write ABCD or E on a form, but the instruction was to give gradings and these could not be given in any meaningful way. The instruction had been given in November but the gradings required had to be based on assessments made months earlier. Rightly or wrongly, each of the teachers had changed to a non-grading approach early in the year and had planned all their year's curriculum, testing and assessment on that basis. In most cases the work on which their assessments had been based had been given back to their students. The records they had kept were structured on an entirely different assessment method. The difficulty was further compounded by the fact that the descriptive assessments and the teachers' records were based not merely on tests or marked assignments but also took into account work in class, oral work, projects, reliability, understanding of concepts, attitude, behaviour, application, self improvement and so on.

By comparison with this, the framework of the Principal's directive (A — outstanding, B — very good, C — good, D — satisfactory and E — not satisfactory) was vague and inadequate. This had been amply illustrated by Mr. Gillett's answers to questions about the meaning of letter grades.

On the question of what penalty to impose if the teachers were found guilty, Mr. Impey submitted that in seeking to implement non-graded assessment the teachers had acted responsibly. They were following a system supported by the Curriculum Advisory Board between 1968 and 1974, and by the Director of Secondary Education,

Mr. R. Reed. No departmental memorandum had ever contradicted this. They had not avoided assessing, but had provided detailed records of immeasurably more use than a letter grade on a mark sheet.

Mr. Impey pointed out that although assessment had been a frequent topic of debate in schools over many years, no dispute about it had ever been referred to the Tribunal. It was a dispute that should have been resolved at the school level.

Within the school, moreover, all democratically made decisions on assessment had supported the descriptive method. Yet the Principal consistently ignored staff views on the matter. In fact his attitude had been entirely reactionary and unreasonable.

The Department, for its part, had chosen to ignore the dispute for two years, despite a number of reports from the Principal, and despite repeated attempts by the VSTA to have the dispute resolved.

Mr. Dane, for the Education Department, started by saying that the Department was not asking for the teachers to be dismissed, but that they did view the matter very seriously, and wanted a penalty that would deter further breaches and have a general flow-on effect on other members of the service.

The Department's case, he said, rested on these points:

- The instruction was lawful and it was not obeyed.
- The question of its meaningfulness was irrelevant.
- No approaches had been made by any individual to the Principal or the Department complaining about the instruction.
- None of the arguments put by the defence constituted a defence as in order to do this it had to be established that the Principal had no authority, that the instruction could not be complied with or that the type of assessment suggested was unlawful.
- The question of graded versus non-graded assessment was not on trial.
- The case of non-ability to comply rested on failure to carry out earlier instructions given by the Principal.
- The teachers' offence 'smacked heavily' of dictation from their union.
- The argument of impossibility of compliance was *post facto* i.e. invented after the event.
- It was not universally agreed that grading was meaningless.
- The teachers were selective with regard to what was meaningful and what wasn't, and in many instances they were intellectually dishonest.
- There was a clearly stated staff policy in 1980.
- The Principal had compromised on the matter by allowing descriptive assessments to be sent home to parents.
- The teachers were simply a 'militant group of teachers opposed to the policy set by the Principal'. (Grant and Hannan, 1981, p. 88)

The Tribunal found the charge of failing to obey promptly a lawful instruction proven and fined each teacher \$100, thus supporting the principle of organizational accountability upon which managerial evaluation depends and rejecting the educational and instructional goals on which educational evaluation rests.

Evaluation and Accountability

Accountability can be construed in essentially contrasting terms depending whether it is viewed from an educational or managerial perspective. From the perspective of *educational* evaluation, accountability is clearly concerned with the exploration and justification of educational practice conducted through dialogue and discussion among teachers, pupils and parents within the context of the school community. The context is local; the criteria are related to the cultural, social and educational needs of the community; the comparisons made are those that relate current possibilities and constraints to current practice; the judgements made are those that concern the truth, beauty and justice of educational arrangements within that context. Where the educational arrangements are found wanting those involved in the practice of education in that school are accountable to the school community for the development of improved practices. Where such improvement is hindered by external constraints, school officials are accountable for achieving alterations to those constraints through administrative and political processes. Through such processes of accountability institutions are made accountable to their members.

From the perspective of *managerial* evaluation, accountability is concerned much less with the improvement of complex educational processes within the context of local cultural needs than with increased efficiency and economy in the achievement of a 'best-fit' between the products of schooling and the social, economic and vocational structure of the society at large. Where changes in these structures are mandated by political, legal or economic processes then schools must be managed in ways that meet such changed demands. Instances of such changes are moves towards greater sexual or racial equality, or towards the 'rationalization' of industry or towards the creation or dismantling of the welfare state.

Managerial evaluation is tied to the processes of policy-making undertaken by centralized government. Such policy-making is an essential component of the 'steering capacity' of government. Policy paradigms which result from the processes of policy-making are frequently 'a curious admixture of psychological assumptions, scientific concepts, value commitments, social aspirations, personal beliefs and administrative constraint' (Rein, 1976, p. 103). As far as education is concerned the admixture is represented by an emerging consensus among policy-makers that

- the purpose of schooling is to provide basic and career skills

- basic skills — especially reading and arithmetic — can be defined with precision
- tests can be devised to measure and compare performance
- allowances can be made for compensating the disadvantaged
- cost efficient education can be achieved through evaluation and accounting systems
- legal and bureaucratic rules are preferable to the exercise of judgement in the achievement of such aims (Wise, 1979, p. 56).

Managerial evaluation is, therefore, directly tied into processes of centralized policy formation and implementation. Regrettably, as Wise points out 'what is missing from this set of assumptions about education is any reference to the *process* of education — to how educational practice actually affects the child. The practice of education is thus to be altered without an understanding of how education occurs' (1979, p. 56).

Thus the conflict between educational evaluation and managerial evaluation is representative of the more general conflict between educational policy-making and educational practice. Where managerial evaluation is used as a mechanism of control directed towards alteration in educational practice severe consequences can arise. Such has been the experience in the United States following the legislative introduction of minimal competency testing legislation (in thirty-three states by mid-1978); accountability laws (at least seventy-three laws by 1974); and accompanying school finance laws, federal legislation (over 1000 pages in 1977) and federal court decisions (over 1200 cases between 1966 and 1970) (see Wise, 1979, p. 2). The result has been the emergence of 'governance by testing' (Shine and Goldman, 1980).

The contrast between the educational view of evaluation and the managerial view illustrates the tension between community oriented school-based educational practice and the legislative models of educational accountability held by central government. The direct effect of this tension between community-based models of educational evaluation and government directed models of managerial evaluation on the evaluation systems of schools is considerable. Clearly, as our case study showed, the tension produces difficult situations for both principals and teachers. The tension is not eased by the conflict between the managerial and advisory functions of education departments, nor is the tension reduced by the conflict between the political adoption of managerial models of accountability and the professionals' preference for educational models of evaluation.

Educational and Managerial Evaluation in Schools

The previous discussion has attempted to give meaning to the distinction between educational and managerial systems of evaluation in schools. It has also attempted to distinguish between the contexts and the criteria that support

each system. In so doing it has suggested ways in which comparisons and judgements between the alternatives might be made. Fairly obviously this paper supports the adoption of educational rather than managerial criteria in this aspect as in other aspects of the school's functioning. It also makes the judgement that the alliance of managerial evaluation techniques with centralized and, to use Wise's term, 'hyperationalized' systems of policy formation and control is unlikely to enhance the quality of learning and teaching in schools. What is more likely to succeed is the use of evaluation techniques involving qualitative, illuminative and collaborative observation, leading to the critical examination and development of teaching and learning processes in schools.

Indeed, techniques of curricular and pedagogical evaluation can be supported in their educational ends only by school evaluation systems based upon qualitative, process oriented techniques. On the other hand, managerial evaluation systems are likely, through their crude partiality and their concerns with social control processes, to confound and inhibit the educational tasks of schools. In particular an exclusive reliance on summative, norm-referenced, performance oriented and standardized tests precludes a fuller understanding of teaching and learning processes. Such understanding might better be gained from the use of formative, criterion-referenced, diagnostic, process oriented evaluations which, in conjunction with collaborative observational techniques, would contribute to a continuing dialogue of educational improvement.

References

- BATES, R.J. (1977) 'On the evaluation of pupils', in *Delta 20*, reprinted in BATES, R. and CODD, J. (Eds) (1980) *Theory and Practice in New Zealand Education*, Palmerston North, Dunmore Press, pp. 199-204.
- BATES, R.J. (Ed.) (1981) *Organisational Evaluation in Schools*, Geelong, Australia, Deakin University Press.
- BERLINER, D. (1980) Seminar on educational evaluation, Deakin University, School of Education.
- BERNSTEIN, B. (1973) *Classroom Pedagogy — Visible and Invisible*, Paris, OECD (CERI).
- BERNSTEIN, B. (1975) *Class, Codes and Control: Towards a Theory of Educational Transmission*, Vol. 3, London, Routledge and Kegan Paul.
- BOURKE, S.F. and KEEVES, J.P. (1977) *Australian Studies in School Performance*, Vol. 3., *The Mastery of Literacy and Numeracy: Final Report*, Canberra, AGPS.
- BROADFOOT, P. (1979) *Assessment, Schools and Society*, London, Methuen.
- BROADFOOT, P. (1980) 'Accountability, social forms and system control: A comparative perspective', paper presented to the Conference of the Origins of Educational Systems, International Sociology Association, Paris, August.
- BURKE, F.G. (1980) 'Testing in New Jersey', in *Educational Leadership*, 38, 3, pp. 199-200.
- CARNOY, M. (1977) *Education and Employment. A Critical Appraisal*, Paris, International Institute for Educational Planning.
- CODD, J.A. (1981) The clinical interview, *Curriculum Studies*, 13(2), 145-150.
- COMMITTEE OF INQUIRY INTO EDUCATION AND TRAINING (1979) *Education, Training*

- and *Employment*, Vol. 1, Canberra, AGPS, February (The WILLIAMS Report).
- COOMBS, P.H. (1968) *The World Education Crisis: A Systems Analysis*, New York, Oxford University Press.
- DORE, R. (1976) *The Diploma Disease: Education, Qualification and Development*, Berkeley, University of California Press.
- DURKHEIM, E. (1977) *The Evolution of Educational Thought*, trans. by PETER COLLINS, London, Routledge and Kegan Paul.
- GALTON, M. (1980) 'Accountability and the teacher: Some moves and counter-moves', in RICHARDS, C. (Ed.), *Power and the Curriculum: Issues in Curriculum Studies*, Driffield, Nafferton Books, pp. 135-46.
- GRANT, J. and HANNAN, B. (1981) 'Principle vs principal', in *Secondary Teacher*, 5, pp. 14-24.
- HARGREAVES, D.H. (1967) *Social Relations in a Secondary School*, London, Routledge and Kegan Paul.
- HARGREAVES, D.H. (1976) 'Reactions to labelling', in HAMMERSLEY, M. and WOODS, P. (Eds), *The Process of Schooling*, London, Routledge and Kegan Paul.
- HAWES, G.R. (1974) Managing open education: Testing, evaluation and accountability: Special Report *Nation's Schools*, 93(6), 33-47.
- HEIN, G.E. (1975) 'Standardised testing: Reform is not enough!' in COHEN, M.D. (Ed.), *Testing and Evaluation: New Views*, Washington, Association for Childhood Education International, pp. 27-31.
- JACKSON, P.W. (1968) *Life in Classrooms*, New York, Holt, Rinehart and Winston.
- KATZ, M.B. (1975) *Class, Bureaucracy and Schools*, New York, Praeger.
- KAY, B. (1977) 'Monitoring: Purpose, desirability and scope', in SUMNER, R. (Ed.), *Monitoring National Standards of Attainment in Schools*, Slough, National Foundation for Educational Research, pp. 14-26.
- LACEY, C. (1970) *Hightown Grammar: The School As a Social System*, Manchester, The University Press, Manchester.
- LITTLE, G. and RANSLEY, W. (1979) 'Testing standards in basic skills', in *Education News*, 16, 10, pp. 20-4.
- MUSGRAVE, P.W. (1979) *Society and the Curriculum in Australia*, Sydney, Allen and Unwin.
- PARSONS, T. (1959) 'The school class as a social system: Some of its functions in American society', in *Harvard Educational Review*, 29, 4, pp. 297-318.
- REIN, M. (1976) *Social Science and Public Policy*, Harmondsworth, Penguin Books.
- ROSENTHAL, R. and JACOBSEN, L. (1968) *Pygmalion in the Classroom*, New York, Holt, Rinehart and Winston.
- SHARP, R. and GREEN, A. (1975) *Education and Social Control*, London, Routledge and Kegan Paul.
- SHINE, W.A. and GOLDMAN, N. (1980) 'Governance by testing in New Jersey', in *Educational Leadership*, 38, 3, pp. 197-8 and 201.
- SPRING, J. (1980) *Educating the Worker-Citizen*, London, Longman.
- STODOLSKY, S.S. (1975) 'What tests do and don't do', in COHEN, M.D. (Ed.), *Testing and Evaluation: New Views*, Washington, Association for Childhood Education International, pp. 13-17.
- TYACK, D. (1974) *The One Best System: A History of American Urban Education*, Cambridge, Mass., Harvard University Press.
- WALLER, W. (1967) *The Sociology of Teaching*, New York, Wiley.
- WILLIS, P. (1977) *Learning to Labour: How Working Class Kids Get Working Class Jobs*, Farnborough, Saxon House.
- WISE, A.E. (1979) *Legislated Learning: The Bureaucratisation of the American Classroom*, Berkeley, University of California Press.

Source: R. J. Bates, 'Educational versus managerial evaluation in schools', in P. Broadfoot (ed.), *Selection, Certification and Control: Social Issues in Educational Assessment*, Falmer Press, Lewes, Sussex, 1984, pp. 127-43.

2

Education in theory, schooling in practice: The experience of the last hundred years

Brian Simon

My topic is the relation between theory and practice in education—a central issue which requires elucidation if significant advances are to be made in the practice of schooling. My interest in this topic has been fuelled over the past few years by participation in a carefully designed observational research project into what actually goes on in primary school classrooms in England—an issue that, perhaps surprisingly, became highly politicized during the last ten years. One thing that emerged very clearly from our findings was the wide gap between what one might call prescriptive theory as defined and formulated in the Plowden Report of 1967 (*Children and Their Primary Schools*), and actual classroom practice itself as implemented by teachers in the schools. So extreme a dissonance seems, to put it mildly, hardly helpful to the teachers, the children, or to educational outcomes generally.¹

One characteristic of the present, or recent, phase is a proliferation of theories about education which often derive from the different disciplines now embedded in the educational arena in a new way. Psychology has, of course, always been present, but not in so many guises, some contradictory, as today; historical studies have for long made their own contribution to understanding; but both of these have been somewhat pressed into the background by two newcomers. One is philosophy, in what its proponents describe as a 'revolutionary' guise; not at all revolutionary in relation to education—it should be added—only at odds with earlier philosophical propositions. The other acquisition, barely present before the war, has been sociology. Born of several founding fathers, seeking to attain all the attributes of a science, sociology has mushroomed of late into a fluster of contending 'schools', each ready to pronounce on key educational issues in one way or another.

The resulting surveys, research findings, theses, publications,

disputations, represent a proliferation of discussion about education which may often be more directly related to the concerns of the particular discipline involved than to actualities and problems in the educational arena. It is clear enough that it would today be out of the question to conduct a social survey which excludes education altogether, as an American author recently did of Britain between the wars, on the grounds that there is nothing to be said about it. On the contrary, in Britain, the United States, and I believe in Australia also, education has been recognized as a very arrow-point of social policy and practice, not only by theoreticians of various kinds but also by politicians of all colours.

As for the average educationist or teacher, hard at work in a tough, demanding and exacting field, he or she has an embarrassment of riches on which to draw in terms of theory—anyway, if sheer bulk and variety is at issue. Coming from such different quarters, the analyses, interpretations, advice offered can hardly be other than contradictory. But apparently it is up to those practising education to set aside large tracts of time, first to absorb the offerings, then to perform the onerous task of co-ordination and extraction of the relevant kernel for use on the everyday plane. It is an odd supposition, to say the least, especially when it is held that all this theorizing casts a disciplined, if not rigorously scientific, light on education for the first time. Science and spawning verbiage do not easily go together. Elegance is more usually the accepted characteristic.

In tackling this topic I intend to adopt an historical approach, and to look in turn at several moments during my chosen century. While it is certainly true that there are differences in educational developments among the different States in Australia, and between Australia generally and Britain, nevertheless there is a certain commonality of development across the country as a whole, while the general pattern of educational and social change over the last century parallels in its key features changes in Britain. If I take my example from Britain, then, the points made should be at least marginally relevant to the Australian situation.

First, I would like to turn back to the two opening decades when measures to establish a universal, compulsory system of elementary schooling were well in train—that is, to the 1880s and 1890s. What was the relation between theory and developing practice as the first national system of schools was brought into action?

Then I propose to look at the decades between the two world wars which punctuated the first half of the twentieth century. By this time

the elementary school system in Britain had been battered down below stairs, as it were, a separate and parallel, publicly-maintained, secondary system having been launched in 1902 on the traditional pattern—which has yet to be altogether outgrown. Moreover, advance was continually arrested by heavy cuts in expenditure mistakenly dubbed 'economies'—a custom recently revived in Britain at least. Nevertheless, the schools remained at work and there was much discussion of education—its proper aims, suitable methods and so on. What were the specific characteristics of this period, and what new influences were operating?

My third period, once more twenty years on, is that from which we have just emerged, 1960 to 1980, when there have certainly been important developments on the ground and a great deal of theorizing about education—but when the relation between the two is by no means easy to define. Concentration on the details of classroom practice in the research project mentioned earlier has led me to see quite a number of topics in a new way, and to appreciate more fully why teachers so often look askance at the research results thrust under their noses, let alone the merely verbal logic of some theoreticians. And, as I said earlier, this is largely responsible for suggesting my present theme.

The Late Nineteenth Century

May I turn now to 1880, the initial year of the century under review: in effect the year when schooling in England first became compulsory up to the age of ten, a somewhat similar situation to that obtaining in Australia. Developments in the wake of legislation enabling the direct election of local school boards, to supervise and administer the building up of a national system of elementary schooling, made this step possible. In 1870 there had been just over one million children in school. By 1896 there were four and a half million and the board schools had come to the forefront, *vis-à-vis* those run by the Church. In particular, the industrial cities of the midlands and north took a pride in the new buildings provided—large, many-storied, spacious, now providing (for the first time) separate rooms for the new system of class teaching; and moreover incorporating a new standard of ventilation and ancillary provision in terms of cloakrooms, equipment, playgrounds, giving them a character all their own. This was the material context within which the battle for attendance was won: school life was to be twice lengthened in the 1890s—the leaving age being raised to 11 in 1893 and to 12 in 1899.

Once a high proportion of children aged 5 to 13 were squarely within the school system, a new perspective opened. With the schooling of a whole generation actively in progress, a new body of teachers came into being required to act in rapidly changing conditions—a growing core of whom were specially trained. What were the ideas to the fore at this period, or brought to bear on those at work in, or for, the new national system?

Perhaps the most striking feature, to the modern eye, was the ready supposition that the ordinary child was capable of intellectual development, and that this could be brought about by sound and thorough teaching in school. In later years, as we know to our cost, this conception was by no means generally accepted. Rather the opposite. What prompted or underlay it in the 1880s?

At a time when hundreds of teachers were facing a multitude of new tasks, as elementary schools developed rapidly, it was possible to depend on a generally accepted psychological theory of learning—associationism. Foreshadowed by John Locke, spelt out in 1749 in David Hartley's *Observations on Man*, popularized by Joseph Priestley, then in due course by James Mill and later John Stuart Mill, it was also taken up by the early socialists, particularly Robert Owen. In 1879 this theoretical approach found expression, in modified form, in a specifically educational context when Alexander Bain, professor at one of the four Scottish universities, published *Education as a Science*. In the circumstances this was bound to exercise an extensive influence, and it is no surprise to find the views there set out faithfully echoed in the many teaching manuals of the time. Indeed, Bain's book was frequently reprinted in the next ten or twenty years. In particular Bain emphasized the physiological basis of mental functions; as he put it, 'the plastic or acquisitive functions of the brain—the property of cementing the nervous connections that underlie memory, habit and acquired power', in a way that looks forward to present-day findings of neuropsychology. Consider also his definition of education as 'the means of building up the acquired powers of human beings'. This is an uncompromising statement which points to teaching and learning as the means of enhancing human capacity. It is a text very different from that to which a later age has become accustomed—the insistence that a child is born all that he may become; that the most teaching can do is bring to light a given, inborn quota of 'Intelligence'.

In the circumstances it is understandable enough that there should have been a firm intention, on the part of the central authorities at least, to keep schooling for the majority short, in order not to outrun

the intended limits of popular education. But, in the event, this is just what the elementary system did, as what were called higher grade schools and 'higher tops' were established by advanced school boards to take their pupils still further, and as systems for training teachers in colleges and on the job evolved. As a consequence there was resort to legislative and administrative means of control but, meanwhile, for a couple of decades or more, a bracing, optimistic outlook held sway which can be traced in the many manuals on the conduct of teaching, particularly those published in the 1890s.

What was the positive core of the pedagogy that sustained and promoted the experience, and extension, of elementary schooling? Central to Bain's approach, as has been seen, is the concept of development, of growth. Moreover although Bain recognizes differences in educability (or what he calls 'facility of acquirement') it is the foundation of his thought that human beings are essentially educable. 'Every impression made upon us', he writes, 'if sufficient to awaken consciousness at the time, has a certain permanence', persisting after the original ceases to work and open to restoration as a 'remembered impression'. Such impressions, or ideas, can be 'fixed' by practice or exercise. The art of education consists in 'economising the plastic power of 'the human system'. We need to search out 'all the known conditions that favour or impede the plastic growth of the system'. It is in this context that particular attention is paid to memory, or what is called 'the retentive faculty' and its cultivation. For on it rests the possibility of mental growth, or the acquirement of what Bain calls 'capabilities not given by nature'.

As presented by Bain, the classic theory of associationism extended beyond formation of the intellect and acquisition of knowledge to the formation of character and the acquisition of attitudes and habits: as indeed it had done throughout the period of the Enlightenment and in the work, for instance, of Richard Lovell Edgeworth, who had sought to put into practice with individual children his own development of effective pedagogical means in the early nineteenth century. Now these conceptions were to be generally applied, and one thing that strikes the modern reader is the extent to which responsibility accrues to the teacher for apparent failures or inadequacies on the part of the child. This follows logically from what is essentially a positive, optimistic theory in its appreciation of the powers of teaching and learning.

May I give some examples of the teachers' manuals already referred to which were published in a constant stream and which passed these ideas on to students in training? Thus Joseph Landon, vice

principal and former 'master of method' (as he is described) at Saltley Training College, published in 1894 *The Principles and Practice of Teaching and Class Management*. This was also the subject of one of the 'Royal Handbooks for Pupil-Teachers' published by Nelson of Edinburgh in 1895, by a former assistant at the Edinburgh Normal School, J. Gunn. In 1896 *A New Manual of Method* appeared by A. H. Garlick, head of the Woolwich Pupil-Teacher Centre, which reached a fifth edition by 1901—a guide aspiring to be 'comprehensive in its range, practical in its nature, and modern in its methods'. Another popular manual was *The Art of Teaching*, published in 1898 by David Salmon, Principal of Swansea Training College. Although it seems that Victoria did not experience turbulent advances as in England in the 1890s, these years saw the publication, for instance, of F. J. Gladman's *School Method*, which reflected this outlook. Headmaster of the Model and Practising Schools at Borough Road, later 'normal master' at the Borough Road Training College in London, Gladman then moved to the Central Training Institution here at Melbourne. All these manuals depend in large part on the precepts advanced by Bain as the spine of an ordered system of pedagogy.

Garlick, for instance, presents a whole list of characteristic behaviours as remediable, provided that the teacher understands how to tackle the problem—whether it be inattention, laziness, stupidity, truancy, or whatever else. Such forms of behaviour may be the result of ineffective action or understanding on the part of the teacher. Thus, it is said, 'To fix and maintain the attention there are certain laws of the mind which must be obeyed, and any violation of these laws is sure to be followed by more or less inattention'. Similarly ignorance or indifference to the elementary principles of child life is likely to generate laziness. Should the task set make little demand on 'child activity' then 'there is a tendency to despise and reject it', and sit back. Likewise stupidity, if sometimes a gift from parents, may also be 'bestowed by the teacher' if he is 'unmethodical or not properly acquainted with his subject', careless or inaccurate in describing or informing, a poor questioner, or harsh or cruel in his manner so as to 'chill or freeze the efforts of the timid, the weak, or the shy'.

More specifically, associationist theory provides an explanation of habit formation which was seen as the key to discipline and character formation, no less than intellectual acquirement, and underlay much methodological advice. Joseph Landon, of Saltley Training College, attributes to the science of education 'the foundation truths' on which

all practice must rest, and goes on to deal with oral teaching and lesson procedures before turning to the details of questioning, illustrations, the lecture, problems of class management and, finally, the teaching of specific subjects; that is, in particular, reading, spelling, writing, arithmetic, but also the secondary or so-called 'class' subjects of history and geography, English and elementary science.

It is germane to note here how much attention is paid to the matter of class management—in a systematic way which has long passed out of fashion so far as theory is concerned. Yet the problems of managing a class remain, for all teachers (as we certainly found on the primary school research project mentioned); and to ignore them, or sweep them under the carpet, is to make a false assessment of the teaching and learning situation which may well be one of the most potent causes of the gulf between theory and practice.

I have provided no more than a brief and partial excursion into the closing years of the nineteenth century when, with the aid of a long-standing psychological approach—deriving it might be said, from the 'Enlightenment' years, but later brought specifically into relation with educational tasks—the emergent teaching profession built up a coherent system of pedagogy, even if the word was rarely used in Britain. This was constructed at a time when there were no effective brakes to prevent the upthrust of the elementary schools, however hard the system of payment by results might weigh on and distort teaching. In all the circumstances, advance was much faster and went much further than had been envisaged.

During these twenty years it was not on advice from universities—bar those in Scotland with an atmosphere all their own—nor by following the example of old established grammar schools, that the national elementary school system was constructed. Practical guidance had come from the training colleges for teachers and pupil-teacher centres specifically serving the elementary system. There was an optimistic theory of learning on which to draw, to build up a serviceable system of pedagogy, while in practice the keynote of the period was expansion, development—more and more children in school, a rising leaving age, a firm push to provide higher classes, the growth of the teacher training element. True, these features were characteristic of the industrial towns and other centres of population. There was a large gap between the best of the new board schools and relics of the old in slums or all-age village schools still under the aegis of the Church. Nevertheless a new base had been laid for educational developments.

When one is making a case, in the attempt to amend the balance of

interpretation, it is easy to go too far. But it is germane to note that a current history of psychology virtually dismisses Bain's *Education as a Science*—whose detailed coverage conveyed to innumerable readers what they most needed—describing it as 'particularly long-winded and tedious'.

Ironically enough this assessment is by L. S. Hearnshaw, whose latest book is a biography of Cyril Burt which is forced, in the face of inescapable evidence, to write off as tainted the work of a most highly-regarded psychologist who, in practice, led teachers in the wrong direction for most of the twentieth century. It is hardly necessary to add that the intervening guide in the 1920s and 1930s, *Education, its Data and First Principles*, by Percy Nunn, widely disseminated to propagate the 'New Education', fails even to mention, let alone recognize, the value of Alexander Bain in the ideas he promoted.

It is no part of my intention to deny the immense difficulties faced by teachers in the closing years of the nineteenth century. Classes could be enormous, premises overcrowded, resources often meagre in an enterprise conducted cheaply enough. Many teachers were unqualified, if not little more than children learning on the job, and over all hung the restrictive influence of a system whereby salary level was dependent on the number of passes obtained by pupils on precisely defined 'standards' in the basic subjects. The system of 'payment by results', though phased out in the 1890s, necessarily confined the curriculum, and led to drill, drudgery, and a strong emphasis on rote learning—which indeed persisted long after the current cause had been removed. All this has frequently been emphasized as the hallmark of the period, quite often in surveys written from the standpoint of the 'New Education'.

But such accounts by no means cover a scene which was informed by a positive pedagogy stressing educability, the 'plasticity' of the human brain, the power of education, systematically conducted in the light of psychological understanding, to *develop* capacity. One swallow does not make a summer but Stuart Macintyre, historian at the University of Melbourne, in a recent book (*A Proletarian Science*) cites the case of T. A. Jackson, a well-known autodidact and popular socialist lecturer in Britain throughout the first half of this century. Jackson was the son of a compositor of radical views, and was educated at a local elementary or board school in London. Macintyre particularly mentions the encouragement given to the boy by his teachers, especially in his extra-curricular interests in a range of

subjects, notably in science and Greek mythology. Jackson was born in 1879; he would have been at school from 1884 (aged 5) to 1893 (if he stayed to 14). He therefore spans this period. I have seen audiences ranging from high-level theoretical physicists to ordinary workers listening open-mouthed to Tommy Jackson, as he used to be known. We should not, perhaps, underestimate what was achieved in the elementary school at that period.

That students in training were advised to pay close attention to discipline, to maintaining their authority as teachers leaving nothing 'to the caprice of the children'—as Salmon put it in the 1890s—is understandable in the circumstances, especially if one takes into account that the system itself had been introduced, in part at least, as a social-disciplinary measure shot through with a class morality. All the more remarkable, it might seem, that it was from within precisely such a system that a positive pedagogic pattern emerged, one which took intellectual development seriously and sought to further it by promoting systematic teaching and learning to that end.

Before the nineteenth century was out steps had been taken to halt the resulting upthrust of the elementary schools which, it must be remembered, were, in Britain, designed explicitly for the working class, as laid down in the 1870 Act. Indeed Draconian measures were taken from 1899 to 1904 which destroyed the school boards, introduced a new administrative discipline and pattern and emphasized the lower status of the elementary schools. No more higher classes reaching up beyond the leaving age to the secondary level were permitted. Instead there was interpolated a separate system of schools giving what we now term a 'secondary' education, charging fees, and taking children from 5 or 7 in parallel with the elementary system. Into these a small percentage of children from the elementary system were allowed to step, sideways, at the age of 11. So opened a new phase. But whatever changes were to supervene, the basic classroom situation remained constant and the core of the pedagogy, consolidated at a time when the first national system of schools was born and teaching became a profession in the modern sense, persisted as an important element of school practice.

The Inter-war Decades

If we turn now to my second period—the inter-war decades of 1920 to 1940—we have a very different picture, at least as far as educational theory is concerned. In place of the optimism of the 1890s we find on the one hand a pervading fatalism—a stress on the limitation of

human powers based on what I have characterized elsewhere as the 'iron laws of psychometry'; on the other hand, seemingly by contrast, the rhetoric of the 'New Education'—or, as it came to be known, 'progressivism'—which operated, as Professor R.J.W. Selleck has pointed out, as a kind of persistent obligato, above and beyond the actual practice embodied in the system of mass schooling. The scene has certainly changed—and rather fundamentally.

The various strands which made up the 'New Education', which began to make an impact at or around the turn of the century, have been extensively analysed by Selleck in his book of that title. There he brings to light the wide variety of approaches subsumed under this heading—classifying the different groupings as the practical educationists, the social reformers, the naturalists, the Herbartians, the scientific educationists and the moral educationists. In tracing the origin of ideas he points specifically to a genesis not from practice in Britain but from overseas. 'The naturalists', he writes, took their two main deities, Pestalozzi and Froebel, from Germany, while lesser gods, such as Montessori and Dewey, came from Italy and the United States. From Germany, too, though via the United States, came the Herbartians. From the United States for the most part, though originally from Germany, came the 'scientific educationists'. Nor is Sweden to be forgotten as the source of that approach to manual training known as 'Sloyd', and of Ling's gymnastics.

It is clearly not possible, in a single lecture, to do more than point to this eruption of ideas at this specific time. Duly modified, or transmuted for conveyance to an English audience by English disciples, these ideas were sometimes mingled and amalgamated and no doubt to some extent affected practice. But, says Selleck, 'they were never submitted to a searching analysis', so that contradictions and ambiguities prevailed within what passed as the New Education movement. This left space for practising teachers to manoeuvre, to select relevant aspects from the spectrum on offer to fill out gaps or expand aspects of the workaday pedagogy evolved for the elementary school.

Here, then, is a certain relation between theory and practice—a certain give and take characteristic of the period when the initial explosive development of elementary education had simmered down. Then came World War I, with its unprecedented horrors and, too, promises of reconstruction at a new level, promises unfulfilled during the next twenty years when retrenchment supervened, together with the depression and that attendant mortgaging of the future inflicted on society when educational expenditure is heavily

cut. Interestingly, it was during this black period that a 'progressivism' emerged to occupy the centre of the picture by 1939.

This process is generally held to have begun with the publication in 1911 of Edmund Holmes's book *What Is and What Might Be*. Insisting that the child should be at the centre of attention, Holmes held up to obloquy the whole tradition of what he called 'Western education', as something imposed upon children requiring 'blind, passive, literal, unintelligent obedience'. This denunciation applied to the traditional secondary school no less than to the mass schools so organized as to drum the elements of literacy and the ruling morality into the prospective adult workers and voters of the twentieth century. By contrast Holmes described the free, joyful activity of individual children, allowed to be themselves, prevailing in a small 'progressive' school in Sussex run by an understanding 'Egeria', which Holmes aptly dubbed Utopia.

Here progressivism was held up as the ideal and, by definition, an ideal is unattainable—something to be striven towards but never reached or realized. There had been danger to the established order and institutions in the power and force of the upthrust from the elementary system up to 1900. There was no danger at all in the possibility that the inter-war elementary school might turn 'progressive'. In any case any such suggestion was likely to be met with strong opposition from teachers immersed in quite another order of activity in the daily affairs of the nation's elementary schools.

In sum, there was something more than an element of unreality in proposals for educational change which presupposed small classes, plenty of equipment, gardens surrounding classrooms, not to mention freedom from the pressure of the 'scholarship' exam at eleven which in England now virtually determined the core of the elementary curriculum. Nevertheless, whenever reforming advice was proffered (as for instance in the various consultative committee reports of these years) it tended to juxtapose 'progressivism' to current practice, to the detriment of the latter. Thus the former was held out as the only desirable way forward. In effect, given that the approved way was out of the question, this was to deprive the elementary school of a perspective.

In fact, practical predominance was achieved during this period by the so-called 'scientific educationists' (whose concept of a science of education contrasted sharply with Bain's). The emergence on to the stage of mental testing can be traced in one of the manuals produced in the first decade of the century, in this case by a professor at one of the new education departments rising to prominence at modern

universities, J. W. Adamson of King's College, London. Entitled *The Practice of Instruction*, this was first published in 1906. Re-issued in 1912, it now carried a largely rewritten section on 'experimental pedagogy' which marked the eruption of a new brand of psychologists into the field. This new section on measuring 'intelligence' incorporated references to works by Spearman, Brown, Burt and other psychometrists. This was, of course, the period when the Eugenics Society, founded in 1907, got under way, propagating the theory (then taken as axiomatic) that heredity is the dominant influence in determining human characteristics, including intellectual powers; an idea which earlier had had little or no impact on educational thinking, and scarcely finds mention in the manuals of the 1890s.

The significance of mental testing, from our point of view, lies in its implicit denial of the power of education to bring about human change—at least in what was defined as the primary mental characteristic—intelligence. This was a fundamental re-orientation of psychological theory, and it operated to place learning theory on the shelf, literally for decades. In Britain this was of profound importance, since it was there, above all, that the theory and practice of mental testing became closely intertwined with the whole system of schooling, as it developed in the inter-war period and beyond. The processes of streaming and selection, which dominated the period, were both operated and legitimated by theories derived from mental testing. There was, in this sense during this period, a close link between theory and practice—but at what a cost!

'By intelligence', wrote Cyril Burt in 1933, the 'official' psychologist who most closely influenced developments in Britain at this period,

By intelligence, the psychologist understands *inborn, all-round intellectual ability*. It is inherited, or at least innate, not due to teaching or training; it is intellectual, not emotional or moral, and remains uninfluenced by industry or zeal; it is general, not specific, i.e., it is not limited to any particular kind of work, but enters into all we do or say or think. Of all our mental qualities, it is the most far-reaching; fortunately, it can be measured with accuracy and ease.²

And, related to this, what was the function of the school system? As late as 1950 Burt spelt out the role of education in what he described as the 'ideal community' as follows:

Obviously, in an ideal community, our aim should be to discover what ration of intelligence nature has given to each individual

child at birth, then to provide him with an appropriate education, and finally to guide him into the career for which he seems to have been marked out.³

So, in the name of tailoring education to the needs of the 'individual', as diagnosed by mental testing, children were allocated to specific streams from the age of seven or even earlier, and to different and parallel types of post-primary schools where, in fact, they were still treated (inevitably) in the mass. The key words were concentration and classification, each differentiated group of children to be treated according to predictions about mental development derived from testing—predictions which placed clear limits on what might be achieved. Whereas in the late nineteenth century the stress had been on what children have in common, on the possibility of developing human potential, on the enlarging of the intellect through the acquisition of knowledge and on devising teaching to this end, as the twentieth century gathered way, all the emphasis came to be put on the limitation of human powers, the low or mediocre level of intelligence with which most were endowed, and on the impossibility of improving this. By definition, one half of the child population was, and was bound to be, 'below average'—according to the precepts of the psychometrists. Once it had been the proposition that failure to advance on the part of the pupil reflected adversely on the teacher's technique and command of learning theory. Now the stress was well and truly on pupil deficiency. With the basic elementary school no longer on the upgrade, but battened down below hatches so that it was impossible for the majority to advance beyond a certain point, a theoretical outlook came to the fore to prove that this was just as it should be. That, by nature, most children cannot achieve much academically (or intellectually), so that marking time in school with no perspective up to the age of fourteen is all that is required.

How do we account for the coincidence of the hegemony of mental testing on the one hand, and of the ideas of the 'New Education'—or 'progressivism'—on the other? For these ideas intertwined and reinforced each other in the period under review, and are, for instance, clearly fused in the outlook of Susan Isaacs, an influential educationist in the 1930s, who combined hard-line psychometry with pioneering work in what might be called 'liberationist' theory and practice both at the famous Malting House School at Cambridge and in her well-known books on the social and intellectual development of children. Evidently both sets of ideas and practices stem from the same source—belief in inherent, given or inborn qualities of the child

as determining development. Is it not this kind of reductionist biologism which lies behind both the belief in the determining character of inborn mental powers, and the idea that the function of education must be restricted to the flowering (or maturation) of inborn qualities as the Froebelians claimed—a process which, it came to be held, requires that the teacher stand away from the child in order not to interfere, or distort, his or her natural spontaneous development? Is it any wonder, then, that given the economic and social circumstances of the time, a watered down 'progressivism', as Selleck concludes from another study of the period, 'became the orthodoxy'. 'Most of those who wrote books on education', he writes, 'spoke at conferences, produced official reports or sat on important committees, trained teachers, or contributed to educational journals came to accept progressive views as a basis for their own thinking', adding that 'as the twenties and thirties progressed it became clear that the progressives were dictating the terms of the debate'. By 1939, says Selleck, it could almost be said, and particularly in relation to the primary school, that they constituted the 'intellectual orthodoxy'.⁴

If so, it was an intellectual orthodoxy perfected over and above the school system, where the pedagogy of mass schooling perforce held sway, though now, as compared with the 1890s, having very different perspectives. It was during this period, and as a result of this experience, that the forces matured that were to force a new transformation in the years to come.

The Proliferation of Theory

Turning now to my last period—1960 to 1980—we find, perhaps not surprisingly, an entirely different scene. Even if some of the deficiencies in the last period I have discussed seem now to have been, at least partially, overcome, the increasing sophistication of educational theory meant that it was now in danger of taking off far away from the practice of the schools themselves; indeed much of it must appear unintelligible to practitioners.

It is not difficult to account for the proliferation of theory since 1945, or more particularly, since the 1960s. It is a direct result of the expansion of higher education; and, with this, of the advent of new areas of study or the mushrooming of areas barely established in the 1930s, such as philosophy and sociology. This is the context in which it was decided, back in the 1960s, that a viable theory of education is only to be attained by bringing to bear established disciplines in the academic field. Those in favour of the move, initially the 'new

philosophers', had little difficulty in exposing the educational theory prevalent in colleges of education—which, at the time, were undergoing an enormous expansion—as a mishmash of ambiguity and contradiction. Thus the 'progressivist' invasion, so apparently successful down to 1939 and ready to take over after 1945 had its quietus—up to a point; at the theoretical level, shall we say.

I have already referred to the contemporary scene deriving from these developments, at the start of this paper, so there is no need to elaborate on it now. My main conclusion there was that what is now being produced in this field too often bears little relation to the actual practice of schooling—on the ground, as it were. Of recent years a main focus has been on overarching theories concerned with the relations between education and society—the work of Bourdieu, for instance, or of such social analysts as Bowles and Gintis.⁵ A main conclusion appears to be that education inevitably reflects the interests of one or other of the dominant power groups or classes in society as defined variously by these authors; it is either concerned with the reproduction of the direct needs of capitalist industry or with those existing class and group relations necessary for the maintenance of the status quo. Education as a whole is allotted little or no autonomous power or scope for activity. In either case the educator's task is defined, and indeed determined by, inexorable forces quite outside his control. Such analyses tell us little about the *process* of education and indeed, in some respects, exhibit a fatalism rather similar to that embraced by the psychometrists of the 1930s, if on different grounds.

One difficulty with sociological—or socio-economic—surveys of this kind is that they seem somewhat naively to reflect the prevailing socio-economic atmosphere, so that in this sense they hardly contribute to elucidating what one may refer to as strictly educational issues. For instance, a well known text in the sociology of education published in 1961—*Education, Economy and Society*—reads now as a paean on the crucial importance of educational development and growth as the essential ground both to achieve more rapid economic and technological advance and to produce a more 'egalitarian' or 'meritocratic' society. Its successor, significantly now entitled *Power and Ideology in Education*, appearing in 1977, is basically a set of papers highlighting the ineffectiveness of education to effect any social change—a tergiversation closely paralleled by the fluctuation in the world economy.⁶ For the historian all this provides interesting data on the movement of ideas and their relations to social change. But for the classroom practitioner little of significance remains. In particular, such approaches appear profoundly hostile to learning

theory and its educational implications—that is, to areas of concern which must constitute the heart of any theory of education.

Education as a Process

What can we conclude from all this? In two of my three periods, at least, there seems to have been a close relation between theory and practice. The first of these, 1880–1900, was a period of impressive educational advance. For a whole concatenation of reasons, and from a variety of motives, it was thought that the masses should be educated, or at least schooled—and they were. This whole enterprise was, as it were, powered by an ideology—or theoretical stance—which emphasized the educability of the normal child, a view underpinned by advances in the fields of psychology and physiology relating to human learning. Indeed these views still reflected those of the English philosopher, John Locke, who held that nine parts in ten of what a man becomes is due to his education and upbringing. In this period, then, the view of education as powerfully formative was predominant—an outlook that accorded well with the role of the teacher as then defined, and with classroom practice which is and must be primarily concerned with promoting learning in students, with developing particular skills and abilities, with intellectual, no less than moral and aesthetic development.

Interestingly enough, and appropriately for this lecture, it was precisely this aspect of British developments at this time—developments that were effectively halted and transformed in the early 1900s—that were picked up and reported on at length by the Royal Commission on Technical Education, chaired by Theodore Fink, in the years 1900 and 1901. Not only do the Commissioners quote John Locke, but they see 'the real object of education', as they put it, as 'the development of intelligence, faculty and character'. Full descriptions are given of the most advanced schools and colleges in England, shortly to be brought under the axe. It may be that the years that followed, when under the leadership of a galaxy of State Directors of Education, Australia in her turn moved ahead, are years of contrasting developments in our two countries.

If the 1880s and 1890s stressed the power of education, and hence of environment, in human development, the opposite was the case in the period 1920–40, and later. Now it was the power of hereditary endowment that was claimed as all important; a theory or ideology that tied in closely with what had now become a stratified, rigidly controlled and hierarchical school system that seemed closely to reflect the economic, social and even political stagnation of the

period. Once again, theory and practice seemed closely related—but in a determination to prevent advance and change; in an entirely opposite sense to the 1890s.

This prompts the thought as to whether the time has come when we should break out, once and for all, from the heredity/environment dichotomy which has bedevilled educational thinking—and theory—for so long. One positive feature of the last twenty years has been the increasing understanding of education as a *process*, or series of processes, in the course of which specific abilities and skills are developed in the child through his or her own activity with the guidance and help of adults. This I conceive to be the main significance of the work of Jerome Bruner in the United States, of Jean Piaget, of the late Professors Luria and Leontiev in the Soviet Union—both pupils of the remarkable Lev Vygotski who died young in 1934, but who stressed the formative power of the child's life experiences (and especially schooling) in the development of the higher mental processes, and in particular the role of language in this formation. What is ignored in the mechanistic heredity/environment model is precisely the child's own self-activity through which he or she masters and transforms the surrounding world, and in so doing transforms himself—in short, learns, develops, acquires skills, abilities, an outlook, character, attitudes, autonomy—and all the better with systematic aid from adults, in the family, the school and elsewhere. It is on this *process*, I suggest, that we need to concentrate and focus our attention if we are to bring theory once again into close relation with practice, but now on a new level—with the aim of achieving a new, purposive educational advance, now an urgent necessity (as many think) in view of the implications for education and life styles generally of the micro-electronic revolution, automation and robotics.

The focus, then, for study and research must, I suggest, be on education conceived of as a process. It is agreed that education is not and cannot be an autonomous discipline; that it must rely for its theory on psychology, sociology, philosophy and advances in related areas—and in the 1960s to 1980s this has been the dominant approach. While recognizing the importance of the contributions made in these fields, let us recall that education is in fact a mass social activity and as such is clearly open to investigation; as the Bryce Commission on secondary education noted in the 1890s, in a sense every lesson, or joint activity involving teachers and pupils, partakes of the nature of an experiment founded on certain hypotheses. What is needed is a new kind of operational research, involving both teachers and researchers, which seeks directly to penetrate into,

illuminate, and so to improve the process of education—above all to increase its productivity across the board. Encouragement of school-based curriculum development (as in Victoria) may well prove a good 'way in' to such an approach which itself can generate theory closely related to practice.

What we now need to aim at has, I think, been well expressed by the sociologist, Bourdieu, in a passage analysing how it is that the mass of the children, particularly from the working class, fail to survive a school system whose ethos and procedures are alien to them. What is needed, he says, echoing the great seventeenth-century Moravian educationist Comenius, is 'a rational and really universal pedagogy', which takes 'nothing for granted initially', and is organized 'with the explicit aim of providing all with the means of *acquiring* that which, although appearing as a natural gift, is only *given* to the children of the educated classes . . .' We need to develop, he goes on, 'the most effective methods of transmitting to all the knowledge and the know-how which it demands of all'. Such a pedagogy, he adds, must be 'explicitly and technologically methodological'.⁷

Such a perspective may be controversial, but I suggest it provides an objective which could and should be embraced by any society calling itself democratic that has a concern to maximize the development, intellectual and otherwise, of all its citizens. But there is a further implication—to achieve this we need, I suggest, to shift the emphasis, both in theoretical work and in our practice, from concentration on the *differentiation* of children one from another to focus on what children have in common, their shared characteristics. The former emphasis rose into prominence when the selective function of school systems was dominant, as in the 1930s and later. Today the challenge is to raise educational sights right across the board, and modern psychological thinking stresses that all, except those with pathological disorders, are capable of learning, and that for all, the process of learning is broadly similar. It should, therefore, be possible to winnow out general principles of teaching underlying practice and appropriate for all (though modifying these if necessary to meet specific individual idiosyncracies); and to develop genuinely unified school systems which are a condition for minimizing early differentiation. Contemporary concern for a common or core curriculum, now widely discussed both in Australia and Britain, underlines this necessity.

If we can find the way, and gain the resources necessary, to implement modern knowledge relating to learning and teaching (and that is quite a challenge in the present climate), then we may be on

the edge of a new period which historians of the future will assess if not as a golden age, then at least as one when theory and practice were again unified, on a new level, to the advantage of human development and social change. But no doubt hard battles lie ahead. The historical record clearly shows that there is nothing inevitable about educational advance. Far from progress being linear, advances are more often met by setbacks, by new crises, by ideological and political struggles of all kinds. Our present age is no exception. The chances of success will be the greater, I suggest, the more clearly we recognize the obstacles that must be overcome if we are to turn existing potentialities into reality.

NOTES

¹ Maurice Galton, Brian Simon and Paul Croll, *Inside the Primary Classroom* (London, 1980). See especially chapters 4 and 5.

² Cyril Burt (ed.), *How the Mind Works* (London, 1933), pp. 28-9.

³ Cyril Burt, *Listener*, 16 November 1950.

⁴ R. J. W. Selleck, *English Primary Education and the Progressives, 1914-1939* (London, 1972), pp. 120, 156.

⁵ P. Bourdieu and J.-C. Passeron, *Reproduction in Education, Society and Culture* (London, 1977); S. Bowles and H. Gintis, *Schooling in Capitalist America* (London, 1976).

⁶ The first volume cited (New York, 1961) was edited by A. H. Halsey, Jean Floud and C. A. Anderson; the second (New York 1977) by J. Karabel and A. H. Halsey.

⁷ P. Bourdieu, 'The School as a Conservative Force: Scholastic and Cultural Inequalities', in Roger Dale et al. (ed.), *Schooling and Capitalism* (London, 1976).

Source: B. Simon, 'Education in theory, schooling in practice: The experience of the last hundred years', in S. Murray-Smith (ed.), *Melbourne Studies in Education* 1982, Melbourne University Press, Melbourne, 1983, pp. 86-104.

3

Assessment: A workable democratic method

Bill Hannan

In one way or another, most of us accept that the methods we use to assess students have a far-reaching impact both on what we teach and on how we teach. Primary schools usually adapt and order their programs according to assessments of children's progress in reading and number. Regular assessment continues throughout secondary school and eventually determines whether or not a student can stay in the school system at all. In some school systems, assessments decide not only what course you may take but what type of school you may attend. School assessments in other words are not simply professional judgements of the academic work of individuals. Assessments embody and reflect general ideas about the nature of learning and the purpose of schooling. In extreme cases, assessments may not merely impinge on these questions of purpose but actually determine them. IQ testing, for example, can be used to set the pattern of instruction for an entire school system. So too can the systematic use of graded tests in reading and number.

For a democratic school system, it is important to search for a workable, democratic method of assessment. Such a system should be both readily understood by the whole community and contribute to a general rise in our standards of learning. In saying this, I am not trying to denigrate the achievements of a century of public schooling. All I'm saying is that if we can think of ways to improve the basic education of the mass of young people, we ought to try those ways out. Current forms of assessment, I think, are notably undemocratic and could easily be bettered.

I should point out that I am using the term 'assessment' to mean the judgements we make about students' work. I would like to think that those judgements can be separated from judgements about the students themselves, and that they ought to be separated from judgements about the courses we set for them. In other words, we should be spared ambiguities such as: 'Jane does not have the ability to pass this course' which appears to blame Jane without taking the course or the teaching methods to task at all. If courses cannot be done by diligent students then we also need some evaluation of the course and some evaluation of the system that requires such courses before we decide firmly that it's the student's nature — her 'ability' — that is at fault. 'Evaluation' and 'assessment' are often used as synonyms. It might help debate a little if we adopted 'evaluation' as the term for judging whether a course is suitable for a given student, and reserved 'assessment' as the term for the judgements we make about a student's work.

It could help, too, if we distinguished 'assessments', which ought to advance learning in some planned manner, from 'credentialling' or 'certification' which serves narrower social purposes and is clearly subject to more direct social control. Assessment of students during their formal learning, in other words, stands at some sort of mid-point between the processes by which we devise school courses and classwork, and the processes by which we grade the population for further education or for the job market; and it is affected by both processes.

The two most common forms of assessment are graded assessment and descriptive assessment. In a 1980 survey by the Victorian Secondary Teachers Association, three quarters of a sample of Victorian high schools claimed to use either percentage marks or letter grades usually in combination with a descriptive statement. The remaining quarter used descriptive statements only.

The critical difference seen by Victorian teachers between graded and purely descriptive assessments is that the former are inherently competitive. Whilst it can certainly happen that a whole class may be encouraged in their learning by a judicious use of marks out of ten or A's, B's and C's, on a large scale the inbuilt requirement of a graded system to sort students into good, indifferent and bad, or into passes and failures, inevitably asserts itself. Of course, failure is not in itself a bad thing. For most of us it's a fact of everyday life. But when it is happening constantly at every level of schooling by the deliberate application of a teaching method, we must ask whether the method is promoting the greatest possible learning for the largest number of people, and also whether the method is grounded in solid knowledge about the nature of learners and learning. What, in short, does the failure represent? What is it a judgement of? What has the student (as opposed to the course or the teacher or the school system) done to merit this undoubted odium?

None of these questions is easily answered because in practice grading is a rather opaque and cryptic system. The apparently clear-cut, A, B and C doesn't yield much clear-cut information under cross-examination. In fact, I watched the concept being questioned by a barrister early last year when eighteen teachers from Preston Girls High School in Melbourne were being disciplined by a Teachers' Tribunal for refusing to grade their junior students, and the A's, C's and D's proved very hard indeed to explain:

Question from Mr. Impey (for the VSTA); What is an 'A'?

Mr. Gillet (Principal, Preston Girls' H.S.): *It would indicate that a student has a very clear idea of the subject and has ability in that subject. It indicates that the student has ability to handle this and should be encouraged at all costs to continue that subject for quite a lengthy period of time within the school.*

Is that the top 5% or 20% of students in that class?

It does not go on percentages but on ability. You would not say the top 5% get A's.

Taking you back to the directive of 1979 at that stage A's outstanding; A's more than 80%.

If they had a written test they would receive 80%.

What you are saying is the 1979 directive on that A means more than 80%, but in this one it does not necessarily relate to any percentage at all. You do not relate one to the other.

No, I have not related.

'A' means all of these things you have just described. Does it mean that particular student within that particular class?

It refers to a student.

In the context of that class that he is in?

I cannot see what you are getting at.

If Joe Bloggs does well in Form 3A does that A mark his ability in relation to Form 3A or B or C and D?

It indicates his ability to handle the subject and to continue and proceed further.

If you get a student who does not do any work at all but happens to perform well in assignments, does that A tell you anything about it? Would he get an A?

It would depend on the person rating the assignment what they were coming up with there.

Do not these markings, really relate to the person doing the marking as well?

It comes down to who is doing the job.

If you were asked to give somebody an A, B or whatever, are you asking them to do that on the basis of others within the form? It may be largely a migrant form and do they take that difficulty into account?

Where there is a large number of students having problems such as migrant students, the instructions are for discussions with teachers who take that class. You are asking as to whether it should be moderated in relation to that person's ability in that language.

All this comes into whether you give an A or not.

Those are some of the factors.

If it was based on a statewide situation, would it be based according to all the state or compared to the Preston Girls' High School? What is your framework of reference for these gradings?

The co-ordinators are asked to examine the nature of these tests so as to regulate these requirements or assignments. These people are not first year out people, as they have had some experience and they are hopefully in a position to advise staff on such things and the staff are directed to them to seek such advice.

Let us take in middle of the range C, Good. That can involve a whole series of things. It can take into account whether a student has a language problem?

Yes.

Can it take into account behaviour in class?

Behaviour is not necessarily a reflection of their learning and knowledge. This is a separate facet that can be commented on in a different way. Behaviour in class is one way such as their interest and co-operation to supply answers and doing work. I would hope that type of thing would be considered by such a person.

What if you had a student who is academically bright and only hands in one out of three assignments, but what he hands in is brilliant?

If the teacher is going to remark that "All work asked to be submitted had not been submitted", they may not necessarily give a high rating because the youngster has not done what was required.

For all of these letters you put, there is a range of subjective matter for an individual teacher to determine what mark is given. *There is less subjective matters if there are tests, than if there were non-competitive, non-graded assessments.*

To arrive at these marks and for an individual teacher to give an A, B, etc. would involve a high number of subjective factors. *It would involve the knowledge of tests that had been given. It would involve the amount of work a person has done in class.*

Whether they have particular learning difficulties?

That would also have to be considered, such as if the child knows the subject and if the subject has great emphasis on the language, such as humanities. The reading situation is one that would also have to be looked at.

Do they make those assessments within that class that individual is in, or all the classes of that year or the school overall, or the experience as teachers over many schools?

They would refer to the co-ordinators for guidance.

Their own views as to what they should give and what factors that they should take into account, depends on the co-ordinator's ideas as well?

Yes.

It really does not tell you, does it, unless you speak with the co-ordinators and teachers who give the marks, you do not know what factors have been taken into account in giving you that mark do you?

You do know that they have to take those factors into account.

(TTUV News-Secondary Teacher, June 1981, pp.19-20)

Most teachers try to clear up confusion between students' ability and their will to learn by writing a few explanatory comments or calling up the parents to hear comments personally. Quite a few schools also try to separate the two concepts into separate marks, one for achievement and one for effort. But the fact is that unless the two happen to coincide pretty well, the major role must be given to one of the criteria of assessment. What is the status of an assessment that reads E for achievement and B for effort? Or B for achievement and E for effort? In most schools, and in most homes, the achievement mark would carry the day. For the majority of people, grades represent judgements of 'ability'.

The notion of 'ability', embedded in a system that provides for a certain proportion of failures, makes grading a powerful means of sorting students into fairly permanent categories and provides an apparent justification for giving these groups different courses and different qualifications. The grading system is the key device for separating the 'bright' from the 'dull' throughout schooling, and ultimately for excluding large numbers from the school system altogether.

Proponents of purely descriptive assessment hope to remove some of the worse features of built-in failure by doing away with the competitive element of assessment. Basically they maintain that if each student is assessed in the context of his or her own previous performance, the comparative element of assessment will be so reduced that there will be no pass and fail groups. Instead there will be an account of what has actually been achieved, and a structuring of future work to fit the individual. This does not mean, of course, that descriptive assessments must always be flattering. They can be quite severe, but they do have to look for positive signs in the student's learning and try to build on them.

Obviously, I am interpreting here what I see as the underlying rationales of graded and descriptive assessments. In practice, the two could be indistinguishable — everything depends on people's basic ideas about the nature of learning and the purpose of schooling. What I am saying is that by and large, grading systems accept failure and ability as constants whereas descriptive systems try to avoid failure and have a fairly non-committal view of the concept of ability.

The gravest defect of grading as a universal method is that it builds failure into the general education system. In public examinations this characteristic of grading is visible in the form of pre-determined failure rates. Usually about a quarter or a third must fail. Failure in this case has become a mechanistic concept. We are given a statistical definition of standards which cannot be varied by changes in students, teachers, courses or knowledge. It is naive to imagine that roughly similar ideas do not eventually operate right down the line. Our ideas about 'ability' are also fairly mechanistic. It is a rare school that does not set out systematically to establish which children are bright and which are dull — many indeed attempt to establish this before they have taught the children anything at all.

However much such practices might be rationalised for higher education by economic arguments, it is very hard to justify them at all during the phase of universal, and especially of compulsory, schooling. Is it tenable, in the context of democratic ethics, to compel everyone to do something which for a goodly number will be pronounced a failure? Against the clear evidence of failure in the shape of E's, or retardation, or permanent consignment to remedial or other dumbo groups, are the pieties about doing one's best wherever one is really persuasive?

To my mind the only equitable position is to offer a schooling which clearly rewards students for the work they do without bringing it into virtually day by day comparison with the allegedly better or worse efforts of others. The task of teaching, in other words, is to encourage everyone to learn. Assessment schemes should be made to complement, not undermine, that task.

The effect of failure on individuals is admittedly debatable. Commonsense suggests that everyone has to learn to adjust to periodic failure, but that consistent failure eventually puts you off an activity altogether. Educational research hasn't added much to that commonsense view. There are many variables at play, which is one reason why we still use teachers and schools rather than machines and programs. However, if we combine some commonsense psychology with principles of public equity we can probably say that teaching which succeeds in getting learners to work towards clear-cut, attainable and demanding goals is both pedagogically safer and democratically more defensible.

The users of purely descriptive assessment generally have this more democratic aim. Descriptive assessment, however, does not necessarily resolve some of the major problems for which grading can be criticised. It does remove the obvious symbols of competitiveness and of failure, but it is still not clear about its criteria of achievement. Put another way, we could say that a method of assessment should indicate what theories of teaching underlie it. In a way, graded assessments do this because they rest on very widely held, if hazy, ideas of 'ability'. The fact that 'ability' mostly has only the tautologous meaning of 'doing well at school' does not worry many people. This ill-defined consensus in turn matches up with the obvious everyday differences and inequalities among people. Hence graded assessments, however unfortunate they may be thought to be, are generally accepted as reasonable. Very few people step outside that frame of reference to question the extent to which these concepts — some indefensibly vague, others quite commonplace — should be allowed to influence courses and teaching methods, and even to determine access to public schooling. Unless descriptive assessments step outside that framework, they may be not much better.

The trouble with basing assessments, graded or otherwise, on 'ability' is that the concept itself pre-empt's most other judgements and vitiates most imaginable action to improve performance. Unless 'ability' is treated as something readily altered by work — which is an arguable use of the term — it tends to be a fairly deterministic concept. Whether by nature or nurture, some have it and some don't. But its origins, nature and operation are all mysterious.

Assessment based on unknowns is not much use as a teaching method. What constructive use can a learner make of the assessment that he or she lacks the requisite ability? Such an assessment offers neither an analysis of what went wrong nor a recommendation of what to do next.

To baffle people with tautologies or ambiguities scarcely qualifies as good democracy or good pedagogy. A democratic assessment ought to result from a very clear understanding on everyone's part of what has to be done to obtain a satisfactory assessment. Such a requirement is not met by a statement that you must do work that your teachers or examiners deem to be satisfactory. To achieve clarity, we really have to lay the ground for assessment from the very beginning. We have to make assessment part of the teaching process. In brief, we have to be able to state as unequivocally as possible what work is to be done throughout the course and what criteria will be used to judge whether the work has been done adequately.

For want of a better term, I call this kind of assessment 'goal-based', to indicate that the process of assessment actually begins with the setting of the goals of a given course or unit of a course. Whether these goals are set exclusively by the teacher or by some process of negotiation is not crucial to this debate, though negotiation would obviously impose a greater need for clarity in defining goals.

Let me illustrate this kind of assessment with an example from a Year 12 English course introduced this year by the Victorian Institute of Secondary Education for its HSC examination. The course consists of three units : a compulsory one on Writing and two more chosen from a range of units on language and literature. This is the section (8.1.1, page 29 of the final draft version) on goals in the Writing unit:

The goals of the writing unit comprise four activities. In each case the basis of assessment is completion of the work as prescribed.

1. A journal has been kept throughout the course of the unit.
2. 2500 words of finished writing have been presented by the student. ("Finished" implies that drafting, editing and rewriting have taken place and that the writing presented has gone through that process until the students and teacher find it satisfactory.)
3. An analysis of writing has been undertaken, and the student has indicated an ability to comment on the craft, style and form of writing.
4. The student has developed (a) a statement about the qualities of good writing, and (b) a self-assessment of the student's own writing.

At the end of the unit, which may be concentrated into one term or spread over the year, a satisfactory student (we'll call him Patrick Hogan) would have this assessment as part of his English assessment.

Patrick Hogan has achieved the following goals — He has kept a journal throughout the duration of the Writing unit; 2500 words of finished writing have been presented; an analysis of writing has been undertaken, and the student has indicated an ability to comment on the craft, style and form of a number of samples of writing. Patrick has participated in the development of a statement about the qualities of good writing and assessed his own writing in the light of this.

and so on for the other two units of the course.

Had Patrick been unsatisfactory, his assessment could have read:

He has undertaken an analysis of writing, and indicated an ability to comment on the craft, style and form of a number of samples of writing. He has participated in the development of a statement about the qualities of good writing, and assessed his own writing in the light of this.

He failed to keep a journal throughout the duration of the Writing unit, and to present the required 2500 words of finished writing.

Unlike a good descriptive assessment, which demands some art, this sort of assessment requires little more effort to put down than a grade. Basically the question is simply whether or not the work stipulated in the goals has been done. The key processes of assessment have been absorbed into the design of the course and the progressive working out of the goals.

As far as possible the main weight of this sort of assessment should fall on the goals. The clearer they are — in the sense that everyone knows what must be done to achieve them and accepts that they are attainable — the more clear-cut the assessment becomes. In the example I have quoted the goals are not entirely clear. In the case of 'indicating an ability to comment on ... samples of writing', the authors of the unit (David McRae and myself) stipulate that students who can't do this first time up in the agreed manner be given further teaching and further chances to demonstrate their ability. The other goals, however, are attained largely by working through the prescribed processes: keeping a journal, drafting, re-drafting, editing a given number of words, participating in the development of a statement. And the processes are prescribed for everyone. A gifted writer (of the sort who can write a good piece straight off) would not be credited with having done this course, unless he or she actually worked through the goals as set. The assumption is that the course is worth doing and that gifted and battlers alike will get something out of the doing. Obviously, it is not a unit that anyone at all could walk into and pass just by sticking round — it has the typical Year 12 student in mind and the work is very demanding — but for all that, the principle of assessment is basically to reward for work done irrespective of comparative standards. If you like, it favours the hard workers, among whom it makes few distinctions, rather than the smart idlers. It is true that the formal, final assessment does not offer particular recognition to the smart hard workers but we should remember that it is they who will probably get most personal satisfaction from their work during

the course. The full story of assessment can only be pieced together from all the comments and judgements made in the drafting and re-drafting of many pieces of writing.

Since I last wrote about this kind of assessment (in the *TTUV/News* and *The Secondary Teacher* of February 1981) some sympathetic critics have remarked that writing is a particularly fortuitous example. Perhaps it is. In a writing class, much the same basic tasks can be set for a wide range of students and whether or not they complete the task is quite apparent. Reading, by comparison, is harder to assess. We can't prescribe a certain number of pages to be read without devising some means of ascertaining whether the pages have been understood. In our literature units (in the same VISE Year 12 course, page 75) we have goals such as:

- Complete the reading and other activities prescribed in class and as approved by the consultative group.
- Conduct a sustained analysis of the texts. The analysis should be based on detailed knowledge of the texts and concentrate on explaining the themes and ideas implicit in the works and the intention underlying the structure of the work. (What is it about? Why is it written in that way?)

(page 75)

Elsewhere we define 'a sustained analysis' in these terms:

The analysis may be spoken or written or both. In many cases the students who attend class regularly, participate in class work and lead at least one class session with a prepared paper will satisfy the demand for sustained analysis. Students who do not participate sufficiently (attendance is mandatory in any case) will have to be required to give an individual analysis, with the opportunity to repeat at a later date.

Every possible encouragement should be given to students to formulate their own analyses and to integrate the work into their own contexts and ideas. The assessment therefore must not contain any elements of the teacher judging an analysis to be right or wrong by comparison with the teacher's or with standard critical analyses. What is being sought from the student is detailed knowledge of texts and the capacity to draw from the text issues which the student sees as significant.

(page 59a)

I would not argue that all of this is entirely concrete and unambiguous. Such advanced intellectual activity could never be so. But I do argue that there are notable shifts of emphasis away from deterministic concepts of ability towards a concept that active, disciplined participation in a good course will result in useful learning. If that is so, it is a very good reason to develop good courses and to credit those who agree to participate in them and who honour their agreement. That won't be everyone by any means, but it may be more than we currently credit with learning.

Source: B. Hannan, 'Assessment: A workable democratic method', *English in Australia*, no. 59, March 1982, pp. 11-19.

4

Assessment in practice: Competitive or non-competitive?

Graeme Withers & Greg Cornish

Introduction

Most educators accept that the methods and approaches used in assessing students have a significant impact on how and what is taught and consequently on how, what and when students learn. Assessment is an integral part of the process of schooling and the modes of assessment which are chosen not only reflect our views about learning and schooling but also serve to reinforce the impact of such views on society.

In this paper we draw a fundamental distinction between two principal modes of assessment – competitive and non-competitive – and offer a review of the functions of each mode. Further, we attempt to establish a context for this review by relating the functions of assessment (in each mode) to the processes and structures of schooling, to their impact on learners and teachers and comment on what they reflect about various views of learning and of the purposes of schooling.

'Assessment' in the context of this paper, refers to the process of arriving at a view or judgment of the students' work. It should be distinguished from the reports or statements which result from that process, and from 'evaluation', which we take to be the process of making judgmental statements about educational activities such as courses, programs or some other larger aspect of the educational process.

Working definitions

In current educational debates (and jargon) the pundits distinguish two principal modes of assessment – competitive (or comparative of people) and non-competitive (or descriptive of a person). If we accept this dichotomy, then we should, before we get around to related issues such as the reporting to students, parents, or employers of the results of such assessments, define each type.

Competitive assessment has as its fundamental assumption the idea that any student's performance is best defined or described relative to the achievement of other students, hence the use of 'comparative' as an alternative label. The other students may be classmates in 1983, or 'the class of '63, the like of which is my model of what such students should be' – some internalised standard which a teacher might carry about for the whole length of a professional career. Typically, competitive assessment involves the assignment of integers to student performance – '60 for History, for Jerry, but only 58 for Jenny . . . she's not as good'. This provides a happy ground for staffroom argument – 'I can **demonstrate** that 60 is better than 58', or, 'Ten-point scales are **much** more just than percentages' – all of it, dare we say, equally wrongheaded.

Non-competitive assessment has as its basic concern that only the performance of a particular student is observed, and only his or her most recent performances. The student is not compared with the performances of the 'ideal' (or even the 'other') students, nor with old, very outdated information as to 'aptitude' or 'past achievement' or 'potential' which earlier teachers may have registered on report cards or record sheets. This type of assessment has simply to do with recording and reporting as accurately as the teacher can the precise nature of the student's performance. It may be subjective, using terms such as 'good', 'excellent', 'poor', and no doubt in arriving at these judgments, the teacher will be reaching back to some perception of a standard of excellence (or ineptitude) arrived at in the past. However such lapses have to be forgiven because all assessment is a human and subjective process. The criteria being used as bases for judgment are professional, and involve the training and experience of

teachers, for which society pays them their professional salaries. The criteria are not personal; they are not based on unreliable recollections of the performances of 'the class of '63', or at least we hope not.

Every teacher will incline at various times towards one or the other of these two methods. The issue of this paper is to review the two approaches, according to the implications of each for the purposes of schooling, the nature of the learners and their learning, and the fundamental question (all too often ignored) of what constitutes real skill or knowledge. The reader is being asked:

- What do you conceive to be the purposes of schooling?
- How do your assessment procedures reflect these, taking into account the natures of your students?
- How do you testify to society the degree of skill or knowledge achieved by these people?
- How do you describe such skills or knowledge to those to whom you are, as a professional, reporting?

In proposing the questions and offering research evidence and opinions, this paper urges, none too gently, adoption of non-competitive modes of assessment in schools.

Competitive assessment

The most common type of competitive assessment is 'grading'; that is

providing a numerical mark or letter symbol to summarise a student's progress or achievement within a given period of time (Bellanca and Kirschenbaum 1976).

In external examinations and often in school assessment procedures, the allocation of grades is normative. Students are in competition with each other for a predetermined and limited number of As, Bs and so on. The distribution is controlled in accordance with some prior notion of just

how many As there should be, given the candidate population. This notion in turn derives from an assumption about how widely the intellectual skills represented by an A are likely to occur in that group. Examiners or examining authorities think they know what is normal, based on past experience, and devise and operate the system accordingly. In school contexts, the award of comparison-based marks or grades will rarely be restricted to such a predetermined distribution, but teachers often perceive their assessment task as ranking students in terms of some notion of 'ability'. The sorting or separation of students into certain categories (As, Bs . . . or 90s, 80s) is to a greater or lesser degree 'justified' by the clarity with which the teacher doing the sorting can articulate just what constitutes 'ability'. But some notion of 'normal' behaviour usually survives – failure of some proportion of the candidates in school assessments is still expected. The cry of a headmaster of twenty years ago: 'But you can't pass them **all!**' still echoes on Correction Days (or is it more substantial than an echo?).

The idea that grades represent judgments of 'ability' has serious further consequences.

The notion of 'ability', embedded in a system that provides for a certain proportion of failures, makes grading a powerful means of sorting students into fairly permanent categories and provides an apparent justification for giving these groups different courses and different qualifications. The grading system is the key device for separating the 'bright' from the 'dull' throughout schooling, and ultimately for excluding large numbers from the school system altogether (Hannan 1982:15).

In this context we should note that the impact of grading is not to be recognised only in its most public manifestations – external examination scores. Grading discriminates against individuals (and they the majority of the school population) long before a Year 12 examination season. Arguments in favour of such discriminatory competitive assessment bear a disturbing relationship to what Tawney (1951) describes as the 'tadpole' philosophy – that the unhappy lot of all tadpoles is held by some to be acceptable, even justifiable, because a few tadpoles do eventually become frogs.

In recent years, many schools have sought to lessen or minimise (or dilute) the more odious effects of competitive grading by the inclusion of written comments in students' reports alongside the grades, and of special symbols (usually letters) purporting to report desired attitudinal characteristics such as 'effort' or 'commitment to study' or the like. These are offered as a palliative to parents (or a sop to uneasy teacher consciences, perhaps) in an attempt to be 'positive' or to sweeten the pill of failure. They have their attendant problems:

The fact is that unless the two [achievement mark and attitude symbol] happen to coincide pretty well, the major role must be given to one of the criteria of the assessment. What is the status of an assessment which reads E for achievement and B for effort? Or B for achievement and E for effort? In most schools, and in most homes, the achievement mark would carry the day (Bellanca and Kirschenbaum 1976).

It would be interesting, in this matter, to find out just how seriously the students themselves regard the attitudinal 'measures' and whether the pill of failure is ever sweetened and swallowed. It would also be interesting to see, more often than we do, the set of criteria for 'effort' or 'commitment' that makes clear the difference between an A and a B in this column on the report card.

Occasionally one finds attempts to counter the arguments against grading practices in schools; claims are made that such judgments about students are not made relative to judgments about other students in the same class or school; grading of student performance is assumed to take place with reference to certain specified criteria and levels of performance; such grades are not normative but criterion based. The award of grades is not fixed to a particular distribution, in which all comers are competing for the top spots, but rather in response to student behaviour. If a student's performance satisfies the required criteria then the appropriate grade is awarded. We reject this claim that criterion-referenced assessment is by its very nature non-competitive. It certainly overcomes certain undesirable aspects of raw normative grading, particularly with regard to the reliability of the assessments, but a competitive principle still operates. The grades, whether letters or numbers, are generally taken to be symbolic of the superiority of one student over

another, at least by the students themselves. The report cards are handed out *en masse*. Watch the comparisons taking place. One B to spoil a run of straight As is undoubtedly, to the student, an inferior performance – a sort of failure in the tadpole pond. One's chances of becoming a frog are diminished just that much.

Along with this claim, there is an associated claim which seeks to distinguish for different treatment the compulsory and non-compulsory years of schooling. While such claimants would suggest that, even in the compulsory years of schooling, assessment should be competitive, they would hold more strongly that there is some special virtue in promoting competition as an element of assessment in the non-compulsory years of schooling. The distinction is not acceptable. The students who survive to the non-compulsory years do not lose their rights by virtue of growing older. In fact the reverse: the complex development of human personality during adolescence requires that the assessment procedures or instruments be as precise, as individually-oriented, and as sensitive to human needs as society can make them.

What functions do grades or marks serve?

Evans (1976) distinguishes four functions which grades or marks are presumed to serve: administrative (or selection), guidance, information, and motivation (or discipline). In this review, we will split off the words in brackets to distinguish two additional functions.

Administrative functions

Marks or grades are used in decision making at various times and levels of the schooling process. They are used when a student is transferred from one school to another. They are the principal component of school records (and, in senior years at least, one of the props on which a school's reputation is based, though we could hardly claim this as a 'function'). They are used to determine the composition of classes. Few of these functions are conducted with the active participation of the teachers who made the actual assessments: such decisions are usually made on undiscussed and unreflected-upon assumptions about the meaning of the grades or marks, which are presumed also to be commonly held.

Selection functions

Marks or grades are used to indicate whether a student has passed or failed, or which students should be promoted or required to repeat the grade or course; whether (or rather which) students should matriculate. They are used to judge candidates as suitable for admission to special courses or post-secondary education. They are used by employers in evaluating and selecting prospective employees. In addition to the provisos already noticed under 'Administrative Functions', there are more at work here. By the time many of these selection procedures are carried out, the data are often 'cold' (ie outdated) and, very often too, not comparable. In what sense does a B equal 73 per cent or a B from one school relate to a B from another?

Guidance functions

Marks and grades are used in guidance and counselling to identify areas of special ability (or lack of ability), to make decisions about the advisability of enrolment in particular courses, and to decide the nature of courses in which an individual should be enrolled. Rarely are the marks or grades used as the sole criterion for such decision making, but they are, nevertheless, often the most powerful criterion.

Information functions

Marks and grades are the major means used by schools in transmitting academic and (more recently) attitudinal assessments of student abilities and achievements to the students and their parents. Depending on the school, they focus to varying degrees on achievements, progress, success or failure, intellectual ability, physical prowess, and effort or commitment to study. Mark lists also transmit information to teachers of the students in succeeding years, once the selection or promotion decisions have been taken. Increasingly, but still rarely, the marks and grades come with statements of the criteria which are used to determine them, and which distinguish one 'grade' from another.

Motivation functions

Marks and grades are often used on the assumption that they stimulate students to greater or more directed efforts

in their learning activities. They are used for this purpose in determining eligibility for honours and prizes of various kinds, such as permission to participate in group activities, eligibility to play in the team, or membership of special groups. Their impact in these situations is incalculable.

Discipline functions

One suspects that marks and grades are used, more commonly than teachers would care to admit, as a disciplinary measure against the more refractory students. In the bad old days, to total a student's geography exam result and find that the sum was 49 was a great temptation to favouritism or admonition. How much more powerful is the teacher's hand nowadays, with an A to E scale of 'performance' or 'effort' or 'commitment' or 'class participation' available to deal a few, perhaps small but telling, blows against the back row or the far corner. And how often is a student's whole career decided at a form level 'panel' meeting called to discuss his or her total performance in an atmosphere where the student has no advocate?

How well do grades or marks serve these functions?

In addition to the specific provisos or reservations already expressed, research and intuition have a lot to tell us about the validity and reliability of competitive grading as real assessment of human potential. But before we present the evidence and intuitions, we should perhaps explode one of the myths that 'grades expressed in letters are superior to marks expressed as numbers or percentages'. Many teachers will accept that letter grades are preferable, particularly to marks expressed on a fine scale like percentages. There is perhaps more certainty in the user's mind of the accuracy of the classification when a coarse scale like A to E is used, especially if the scale comes complete with a set of criteria. The certainty, or superiority, lies in reliability – the degree to which the user or reader can be sure that had the same teacher assessed the same student on the same criteria and the same performances, the teacher would next day or on some later occasion have come up with the same grade. However this does not mean that letter grades solve all our measurement problems at a blow. In fact the problem of deciding between one region (say, D) and its fellows (E

and C) becomes significant, not to say crucial. The responsibility of deciding on the difference between a good E and a poor D is one which teachers have long faced, if E is a fail (eg 49 per cent or less) and the other (D) is a pass. But the need to distinguish the difference between a good B and a poor A is new (similarly for F/E, D/C, C/B). Previously the mathematics 'took care' of it. It was assumed that there was nothing mystical about the difference between 89 per cent and 90 percent or 39 per cent and 40 per cent. However, the research evidence suggests that such distinctions between letter grade levels are extremely idiosyncratic of the marker. As noted by Blachford:

The research studies showing marker unreliability are . . . numerous . . . one is led inescapably to two conclusions:

- (i) different raters tend to assign different marks or grades to the same paper;*
- (ii) a single rater tends to assign different marks or grades to the same paper on other occasions.*

*Thus, the same or similar work given an 'A' by a teacher on one occasion may be given a 'B' or 'C' on another occasion by the same teacher. Indeed, Ashburn (1938) concluded that the pass/failure of about 40% of tertiary students depends **not** on what the student knows but on **who** reads the papers, and that the pass/fail decision for about 10% of students depends on **when** the papers are read (Blachford 1981).*

Grades may reduce unreliability, but they do not eradicate it.

Another key idea is that of the predictive power of prior assessment. Although one can point to the fact that a grade based on a student's achievement or performance is a good indicator of subsequent possible or probable levels of performance, such grades are not as good predictors as many of us fondly believe. Typical studies – Evans cites several – indicate that the correlation between grades received in one year and those received in later years is about 0.60. That is, only about one-third of the variation in student performance measured by grades or marks can be accounted for by prior grades or marks.

This phenomenon occurs not only from one year to the next but also across longer spans of a student's academic

life. The complexity of predicting longer-term future performance is instanced by a study by Wilkes and Noble (1972) of candidates for the award of Commonwealth of Australia Secondary Scholarships in 1966 - 1968. The relevant results are summarised in Table 1.

Table 1: Commonwealth Scholarship Candidates, Victoria

	Winners 1966	Non-winners 1966
Matriculation candidates 1968	89.5% (n = 416)	71.2% (n = 331)
Matriculation passes 1968	90.4% (n = 376)	81.6% (n = 270)

As Blachford (1981), among others, has noted, it remains unknown whether many of the 30 per cent of non-winners who did not present for the matriculation examination were capable of success. Of those non-winners who did present it may be said that their chance of success was not much less than those who had won a scholarship by being determined, over four aptitude tests, to be of very superior intellectual calibre. Also one worries slightly about the contribution which not winning a scholarship might have made to student decisions to leave school, in matters of intellectual morale or self esteem. Whatever one's feelings, it is clear that winning a scholarship at Year 10 did not guarantee success at Year 12, and that failing to win one was not a prophecy of certain failure later.

Grades do not appear to be very much better, if better at all, at predicting success in post-secondary education. Correlations reported in research studies tend to be of the order of 0.60 or lower.

Kirschenbaum and others (1971) describe this with an analogy:

If we could fill a large pot with the ingredients needed to predict academic success, grades would leave two-thirds of the pot empty.

That is to say, if we filled a pot with all the ingredients needed to predict post-secondary success, secondary grades would count for only one-third of these ingredients.

Grades in such cases are normally of two origins: the scores obtained in the entrance examination (such as Victoria's Higher School Certificate), and those awarded during the course of post-secondary education, once the student has been selected. Low correlations between HSC scores and subsequent post-secondary scores are partly a function of the restricted range of students who actually proceed to post-secondary studies, and partly a function of the differences between the courses at the two levels.

Another issue in the prediction of academic success, or the interpretation of correlations between selection test success and later-on course success, was identified by West and Slamowicz:

The problem with the use of the correlation coefficient is that it describes the degree of association between the two scores across the whole range of scores sampled, while the cut-off occurs only in one region (the bottom end of that range). The inferential link between the correlation coefficient and the validity of a selection test depends on the assumption that the selection test is an effective predictor in the range within which selection cut-off occurs (West and Slamowicz 1976).

In other words, even quite high correlations between entry and later performance might mask a disturbing degree of unreliability (or inefficiency of selection) at precisely that point – the cut-off point – where the students concerned are most vulnerable. In exposing the degree to which they have mastered the content of their courses, such borderline students who are insecure in their control of test-taking strategies, or who are nervous or unwell, or disturbed by their neighbours in the examination room, are doubly at risk.

Research suggests that while grades or marks may be moderate predictors of future success in later school or post-secondary studies (in the order of 0.60), they apparently bear little or no relation to later occupational success. Hoyt (1966), for example, in a review of some 46 studies, confirmed the very low predictive validity of college and other grades when related to job success; this was shown to be so across a wide range of occupations. Anderson (1932) found a correlation of 0.12 between high school grades and superintendent or principal ratings of job performance.

Kappell (1962) found only slight positive correlations between college grades and final salaries attained by employees of the American Telephone and Telegraph Company. Pallett (1965) found no positive correlation between college grades and eight characteristics of job success in business.

It is relatively easy to explain or identify the likely reasons for such research findings. For example, a perfect correlation (1.0) between entry test success and final post-secondary examination success for a group of candidates might seem to indicate that little or nothing had changed the group in the intervening years. Obviously this example is nonsense, but it does point up the fact that we should expect people to change. That is, we must expect, given the nature of educational processing, correlations which are very much less than perfect. A perfect correlation might also seem to indicate absolute efficiency in the allocation of educational resources, such as post-secondary places: that is, that the 'right' people were selected. But the fact that we cannot achieve it is not to be despised: it is happily accounted for by the fact that we are dealing with human beings, who go off and on the 'academic boil', who change direction, have love affairs, and in general learn more off-campus than they do on it. In the matter of correlation between post-secondary success and occupational success, we should recognise that the nature or content of the curriculum may bear little or no relationship to occupational practice. Perhaps the method of assessment in the post-secondary institution is quite at odds with the way in which an employer will assess success or with the criteria that will later be used in such assessments. But the question remains 'Should there be a relationship?'. This forces us back to some fundamental issues raised previously: What are the purposes of schooling (and hence the assessment of individual success at it)? What constitutes 'good' or useful or purposeful knowledge? We accept that there should be a relationship between post-secondary curriculum and occupational practices, though our view of its details might be different from the average employer asked the same question. But we would point out that the limitations, the less-than-perfect correlations, are explicable and acceptable: that the relationship should be no more than generally positive. No profession, not even medicine, expects 100 per cent success from all candidates admitted to its ranks. The point is pertinently made in an anecdote from a colleague of one of the present authors:

Every time I go to see a doctor, I'm convinced that the complaint I've got was one of the 49 percent of diseases he didn't know about in his final examination.

The distinct impression from the research evidence is that marks, grades, and the typical assessment methods associated with their collection appear to bear a limited positive relationship with later academic or job success and that educational administrators and selectors and employers should accept and recognise such limitations. But to say this is not to justify automatically the collection of grades, or their use as selection devices.

With regard to the guidance functions of marks or grades, the situation is clearly different. Selection is concerned with cut-off points between those who may and those who may not; guidance has as its leading theme the assistance of all students, the interpretation of past and recent achievements to direct further learning or choice of roles (and chances) in the working community. The research cited is also relevant, especially that dealing with academic/job success relationships, but there is another fundamental issue relating to the links between competitive assessment and the guidance function. It is put poignantly by Curwin (1976) in the claim that grades represent global generalisations which 'prove to be more harmful than beneficial in the education process' because 'they allow students, teachers, parents and others to draw **false blanket conclusions**' (emphasis added).

While recognising that this statement is itself a generalisation, and a fairly strenuous one at that, we would point out that there may be an intrinsic conflict between the notion of guidance and the practice of basing it on the results of competitive grading. In its conventional garb, a competitive grading system sets out to sort and rank, applying labels like 'successful' or 'failure'. It engenders an attitude towards learning (in students, teachers, and parents) that failure to obtain a good or pass grade demonstrates that the student has little or no ability in the area tested, and would do best to redirect his or her endeavours. A score of 40 per cent in mathematics is often used, or assumed, to indicate little or no ability when in fact it may represent (or hide) considerable ability in the subject. Next year's maths teacher says (or whispers), 'I don't want failures in my class'. A distraught parent says

to the student, 'All right, you can leave school'. Another parent says, 'See, you'd be better off doing accountancy next year'. A teacher's notions of curriculum relevance, reliability or course appropriateness are disregarded, in the face of the stunning blow of a failure in mathematics.

Not so in the presence of the trained guidance officer who is alive to the subtleties of score and performance interpretation, but how often are such officers present when the crucial conversations take place?

These criticisms apply not just to numerical marks. Those who use letter grades in the absence of any other form of assessment are also locked into vague generalisations about student capabilities. What does a bald B tell a student about his or her progress? It certainly tells nothing definite that can promote change or growth, and to assume that students somehow know which bits of class or exam room behaviour stopped them from getting an A or proved them superior to the Cs is rash indeed. If they do know now, will they remember next February when they come to take another course in the same discipline? Curwin takes the point a stage further:

Few teachers, however, use only grades. Most supply other forms of both formal and informal evaluation and feedback. Without evaluation grades are meaningless, with evaluation they are unnecessary. Obviously educators need to spend more time developing and implementing better feedback and evaluation devices which are more useful to the learner (Curwin 1976:146).

To take it a stage further still: can any realistic guidance be offered to a student who, with the persistent reinforcement of low or failing grades, believes or comes to believe that he or she is basically an unsuccessful person? As Hannan points out:

Of course, failure is not in itself a bad thing. For most of us it's a fact of everyday life. But when it is happening constantly at every level of schooling by the deliberate application of a teaching method, we must ask whether the method is promoting the greatest possible learning for the largest number of people, and also whether the method is grounded in solid knowledge about the nature of learners and learning. What, in short, does the failure represent?

What is it a judgement of? What has the student (as opposed to the course or the teacher or the school system) done to merit this undoubted odium? None of these questions is easily answered because in practice grading is a rather opaque and a cryptic system (Hannan 1982:15).

While respecting the point of view expressed in the latter part of this quotation we should still like to take issue with the first two sentences, by pointing out that failure is not necessarily an intrinsic part of human life but very much a culture-bound matter. For example, in their explication of principles for the development of Indonesian culture and society, the Indonesian Nationalist Party at the time of independence made use of a long-standing cultural tradition, called *gotong rojong*, to enunciate various aspects of what was seen to be sound development of educational policy. The tradition, baldly, is one where plagiarism and cheating, as 'negative' social behaviours, simply do not exist; where such lack of interpersonal co-operation as implied in competitive examination is stigmatised as anti-social; where co-operative learning, to bring **all** members of society to the highest possible level of self and social fulfilment, is what is aimed at. What price failure in this philosophy? The 'failure' is the person possessed of some special skill or piece of knowledge who does not pass this across to his or her classmates so that all may possess it, or at least attempt to.

One final caution on the use of marks or grades obtained by competitive assessment in the guidance function is probably well known but, for the sake of completeness, needs a brief mention. As well as indicating that such awards are only moderate predictors of future success, research evidence suggests that they are unreliable and contain a far greater degree of subjective judgment than is sometimes recognised. We need come no closer to the current debate about assessment than World War I, with research by Starch and Elliot (1912a,b,c) calling into question inter-rater reliability: that is, what happens when different teachers rate or grade the same work. In three simply designed studies, they demonstrated that teachers (of high school English, history and geometry) arrived at markedly different assessments of the same paper in their respective subjects. In each of the three subjects, about 100 trained teachers were asked to mark one paper on a scale of 100 points, with 75 being the first passing mark. In

English the 'best' and 'worst' judgments differed by 39 marks. Critics argued that the nature of English made this result not surprising, given the subjective nature of most English assessment. Such critics were astonished when a greater range – 45 marks – was discovered between the harshest and most lenient markers of the geometry paper, in a subject supposedly more 'objective'. These studies were landmarks in casting doubt on the reliability of grading procedures, demonstrating that variation was not a function of the supposed 'subjectivity' of the subject area but a function of the graders. A later study by Tieg (1952) threw fresh light on this. Here the issue was not inter-rater reliability but mark-remark reliability by the one rater. It was demonstrated that a single teacher, marking the same batch of test papers on two occasions two months apart (again on a 100-point scale), assigned pairs of marks to each paper, which on average, differed by 14 points.

It is perhaps indicative of the inertia affecting educational change that restating research evidence more than half a century old is not a waste of effort. It is probable that most professional educators recognise by now the limited information content of grades and marks. However that these limitations are sometimes neglected is exemplified by all the possible local applications that teachers can make of the following anecdote.

Consider the example of a student teacher who receives a B in student teaching because of a difference in educational philosophy with his or her supervisor. A school might pass this candidate up because of the B even though the candidate's philosophy coincides with that of the school, which has hired an A student with a contrary philosophy.

A good screening device implies the generation of information meaningful to decision making. Data such as strengths, weaknesses, unique characteristics, compatibility with the institution considering the applicant, and other pertinent information must be supplied. Grades alone do not supply the necessary information for a fair assessment. With the inclusion of other data, grades are unnecessary (Curwin 1976:150).

Translate the B in student teaching to a B in some school subject; keep in mind that such grades or marks are still the principal means of reporting student achievement;

calculate the number of such judgments which must be made in any one school term in the schools of just one Australian State: imagine the number of inferences made by different individuals of the meaning or interpretation of just one of those As or Bs or Es; and there is the enormity of the problem. It is not uncommon to hear the officials of an examining authority (overt or covert guardians of relatively cheap, quick, competitive, graded assessments) remark on the need to educate the community about the unreliability or invalidity of grades. While this may be praised as the 'next step' (after fifty plus years) the tragedy persists: the failures remain with the students after the unreliability and the standard errors have been forgotten. We are not decrying the need for information: it is surely a function of any assessment process. But for whom? Who benefits? For what purposes? Once we identify the needs, the audiences, and the priorities – for students more than parents – then the guidance function will be better served.

Then there is motivation. 'For how long would you work at something if the external reinforcement you received, say on a monthly basis, was always negative?' Perhaps we ask too many people, all of them young, with personalities at a crucial stage of development, to continue to work in just such an environment, not just for months but for years, and not just to work, but to show perseverance, commitment and (pity helps them not) enthusiasm. As Evans (1976:47) notes, 'Contrary to popular opinion, there is little evidence that grades supply strong positive motivation for most students'. Chansky (1962) points out that research suggests that downgraded students often continue to fail. (Should 'often' be 'mostly'? One wonders.) He also points out that, although in theory students who receive poor grades should get an opportunity to review their learning and be assessed again to demonstrate their increased proficiency, this seldom happens. The class moves on, whether or not the weakest achieve the base, needed level of competence. As well, those who believe that grades encourage (or coerce) students to better work and more success will derive little encouragement from the following study by Chamberlin and others.

Their intention was to investigate the impact of a rigid college-oriented high school curriculum, including grades, on a student's later success in college. During an eight-year period beginning in 1932,

nearly 1300 students from 'test' high schools were matched with an equivalent sample of other students given a number of educationally important variables. These variables included age, sex, religion, socioeconomic background, and previous grade average.

Teachers and administrators in one group of schools – the experimental group – were allowed almost a free hand in determining how they would develop their college preparatory programs. Many eliminated grading. Schools in the second group – the control group – offered college preparatory programs, that included grading, which were typical of the period.

Three hundred co-operating colleges agreed to accept students from the experimental schools on the basis of recommendations from principals. On the measures employed during college, students from the experimental group performed as well as or better than those in the control group. They earned slightly better grades in all subject fields except foreign languages, and were judged by their new teachers to be more intellectually curious, resourceful, and more objective in their thinking (Chamberlin and others: 1942).

Chamberlin's study, which unfortunately does not seem to have been replicated, demonstrated that grading was not essential to motivate students. On the contrary, the results suggested that grading could be a hindrance to the development of intellectual and personal skills. One may add to these results a number of other studies which indicate or even demonstrate that self-esteem, or in current jargon 'self-concept', is a major influence in motivation to learn and, directly, on the assessments of the success with which the learning has occurred.

Brim and others (1964) reported that students who saw themselves as having high intelligence obtained higher grades than did students with equal 'IQ' scores but who saw themselves as not having high intelligence. More recently, Thompson (1978) has claimed that low-graded students are more likely to be self-deprecatory, to lack clear goals and values, to be vulnerable to disparagement by others, to have immature relationships with parents, and to be more anxious and depressed. Such students, once low-graded, may be subject to different expectations

and to different treatment by a teacher or school in such a way as to reinforce their perception of themselves as having low status.

There would appear, from the work of Bloom (1964 and 1977), to be a positive correlation between a student's attitude to subject or school on the one hand and his or her achievement. Persely (1974) found that low achievers in schools which use grading procedures do not appear to be as positive to schooling nor to relate as well to adults as students who do not receive graded assessments. Yarborough and Johnson (1980) found that students with low IQ scores who are given grade-based assessments become less positive towards school, and do not relate as well to their families as do students of similarly low IQs who do not receive graded assessments.

Some of this evidence may stand in need of confirmation from replicatory studies (preferably not relying on the dubious IQ scores), but its main thrust is clear. Blachford summarises it:

Grades, then, may have an important impact on student social relations, academic self-concept and mental health. Certainly grades are not just reports of student academic performance. Grades teach (Blachford 1981:37).

Precisely what do they teach? It would seem that those who are most in need of positive motivation towards learning are the very ones who are most likely to be negatively motivated by a grade-based assessment. There is also evidence, such as that provided by Polk and Schafer (1972), to support the teacher's intuitions that low-graded or low-marked students will inevitably seek out alternative ways of protecting their self-esteem. The ways are common knowledge to the teaching profession – disruptive classroom behaviours. To say they are common knowledge does not diminish their importance to society at large. Nor does it prove that, if a solution were to be found, it would have been found already. Our suggestion in this first part of the paper is that some elements of the solution have been found, by fifty years of research. If we grade or mark we must do it much more circumspectly than in the past because even now it is probable that we do not know quite what we do to our students by marking their efforts and achievements.

Non-competitive assessment

In the first part of this paper, we described a method, or a clutch of methods, of assessment which shapes or delimits the curriculum. As we suggested, there might sometimes be cogent, or at least mildly defensible, reasons for this occurring. In this second section we describe an approach to assessment which we believe relates more directly and more positively to what are generally accepted to be the purposes of schooling.

The most common forms of non-competitive assessment can be labelled or classified

1. descriptive assessment*
2. goal-based or work-based assessment.

Whether the two are intrinsically different remains to be argued by others – it seems to us doubtful whether a definition of 'descriptive assessment' which is not work-based, or of 'work-based assessment' which is not descriptive beyond letter grades or marks, will hold up. Perhaps the distinction is no more than a matter of emphasis on either the behaviours observed or on the goals or work achieved or completed.

Another broad dichotomy of forms might be labelled

1. student participation in assessment
2. assessment solely by teacher(s).

We have no doubt that the subjectivity which is implied in all assessment is strongly present here. No less than with traditional competitive modes, the students rely on the professional training, maintenance of professional knowledge, standards, enthusiasm and goodwill of the teachers who have them in their care, for fair, accurate and trenchant assessment. What does seem clear to us is that students cannot participate in competitive assessment of their own or other students' performances, but in non-

*It should be pointed out that we are not making the mistake of considering all descriptive assessment statements as by nature 'non-competitive'; merely that all non-competitive assessment statements must be descriptive (with words, not merely numbers).

competitive situations their voices may (and should) be heard.

It must be stressed, moreover, that the research literature is not abundant in dissertations on non-competitive assessment, its features or its effectiveness. This is not to say it is uncommon. Much of the conversation which takes place between teacher and parent on interview nights may be deemed to be non-competitive; verbal descriptions of attitudes, general performance, goal-related performance (in isolation from 'class averages' or percentile ranks) would appear to fit any definition of 'non-competitive'. But as a style, or a practice, or – dare we say it – a criterion underlying a philosophy of schooling, it has not yet attracted the attention accorded the normal curve, the raw score, or the (oddly titled) 'standard deviation' from the mean.

It is not even true to say that the reporting aspect (marks or grades versus words) is the feature that divides competitive from non-competitive assessment. Theoretically it is possible for a teacher to have available for non-competitive award a total of 100 marks for each student's performance during a sequence of instruction (term, course, year). Theoretically it is possible for that award to be made in isolation from or independently of, the judgments that the same teacher is making of the performances of other students in the same class. Hence X who did everything to the expectation of the teacher (and him/herself, if student participation in the judgment is accepted) might be awarded 100, while Y, who slacked, was awarded 75, and Z, who exceeded everyone's wildest expectations, went home with 148. (What would a parent or an employer make of '148%?') But such reports remain generalisations. Non-competitive assessment, if it is to be a purposeful judgment of a student's performance, cannot afford the luxury of such generalisation.

There is, too, the political aspect of assessment which underlies this whole paper. Assessment must complement, augment and clarify the curriculum of a school. A recent discussion of descriptive assessment makes the point clear.

The process of assessment should be consistent with the aims and processes of a curriculum.

Control of assessment is tantamount to control of

the curriculum. It follows as a corollary of the school-based curriculum principle that schools have a significant responsibility... to see that their assessments are credible and meaningful to students, parents, the local community and wider community groups. Their assessment procedures should be open to review (VISE 1981:5).

But what the quotation does not make clear is that we are dealing with more than just school policy and actions, we are dealing with the life-chances, expectations and hopes of the students concerned. In taking account of what a student has actually achieved, non-competitive assessment allows a rational structuring (and negotiation) of the pattern of future work which best suits each student. And more than just 'allowing' for individual differences, this mode of assessment demands that they be formally recognised – a hard and highly political demand of a school, or a system, as such institutions are presently articulated, in this country at least.

So the vocabulary and ideology of non-competitive assessment must be seen as much more elaborate, much less rigid and far more concerned with the destinies of individual humans than the standard dichotomous rhetoric of success or failure which underlies competition in our schools. The vocabulary, unlike that of competitive assessment, is not superficially precise, supported by terms such as 'grade', 'score', 'examination', 'reliability'; it is reflected in remarks that take account of the essential subjectivity of all assessment. It makes use of forms such as 'might have, but didn't', 'tried and succeeded', 'needs to'; it takes account of motivation and attitudes which no simple percentage (not even '148%') will ever convey to the parent reader of the final communiqué on Year 12 English or Year 9 Science. The 'ideology', if such a high-powered term is in fact appropriate, which underlies all this terminology is likewise very different from that which supports (or directs) the competitive forms of assessment. It includes concerns which we might label 'values', 'positive discrimination', 'socialisation', 'class dominance' as well as more traditional aspects of assessment such as performance, experience, aptitude and learning. Any school which adopts non-competitive assessment as a criterion for its curriculum policy should have no doubt that it has made a decision that will have far-reaching philosophical and political impacts on the lives of its students.

Giroux puts the point in a somewhat high-flying way:

If teachers and students are going to examine their own lives in order to understand the ways in which they have been formed within the context of the ideological and material practices that dominate the wider society, . . . radical pedagogy will have to give due weight to the life experiences of those students who are the object of its analysis. For instance, at the level of classroom pedagogy, it is imperative that concepts and tools used in the pursuit of social and self-analysis by students be rooted in the cultural capital that constitutes and mediates their relationship to the world (Giroux 1981:3-26).

Whether you consider your pedagogy 'radical' or not, if it is individual student-centred in the sense alluded to by Giroux, and not whole-class dominated, then competitive practices of sorting and selection are essentially incompatible with or contradictory to it.

Kleinig takes the argument a stage further:

The form taken by [competitive] assessment, in schooling contexts generally, presupposes a rigid teacher/student distinction and a more or less static and asocial view of knowledge (Kleinig 1981:185).

Before we reach a statement of the aims of non-competitive assessment which might support all these claims, we should perhaps take one more side-track: the impact which the assessment policies of schools or school systems might have on the wider societies which support and which are served by such institutions. Kleinig again:

The school . . . is for the most part locked into the dominant status quo. Attitudes are developed in the school which are incompatible with autonomy, knowledge is converted into a scarce resource and material is learned which serves little purpose beyond the stratifying one of examination. Only on a very limited basis is there fostered the co-operative endeavour which is crucial to autonomy and a worthwhile life. Talk of competition fostering creativity, spontaneity, originality, etc. is counteracted by the hard facts of conformity, stereotypes and motivation only by the hope for external reward or

fear of failure. The authority structure of the school and the traditional 'teacher knows best' pattern of schooling ensure that those who get on are generally those who conform, or whose dissent takes a non-threatening form (Kleinig 1981:171).

If it is accepted (as by the present authors) that 'co-operative endeavour' is 'crucial to autonomy and a worthwhile life', then it must be realised that such a tenet is incompatible with a competitive assessment system.

Such an outcome is only likely to be achieved in a learning/assessment system which reaches beyond traditional academic disciplines and modes of assessing 'competence' or 'knowledge', and instead (Giroux again) 'promotes social analysis and points to transformative social action' (1981:23).

Schools are, in fact, active in the shaping or re-shaping of the social and political characteristics of the communities within which they exist. More often than not they actively maintain or reproduce the prevailing distribution of social, cultural and economic rewards: less often they are actively engaged in the promotion of changes to the society which are designed to make the distribution 'fairer' or more equitable. As Apple says:

Schools . . . 'process' both knowledge and people. In essence, the formal and informal knowledge is used as a complex filter to process people, often by class, and, at the same time, different dispositions and values are taught to different school populations, again often by class (and sex and race). In effect, for this more critical tradition, schools latently recreate cultural and economic disparities, though this is certainly not what most school people intend at all (Apple 1978).

The assessment principles which schools incorporate into their curriculum policies are concrete expressions not only of the roles which school communities wish to take in such matters as Apple raises but also of the roles which educational administrators and politicians wish the schools to take. Perhaps schools will claim that it is not their responsibility to eradicate social or economic disparities. And perhaps, school by school, we might agree with them. But our view of their responsibility is that they should provide all students with a curriculum which enables and encourages a critical analysis of the

existing 'disparities' and which does not, by its assessment policy, seek to perpetuate such disparities with notions of inability or failure.

Where, in this argument, is 'reality'? Some kids are able: some are not. Is not the notion of 'inability' merely another way of saying, 'There are qualitative individual differences which even proponents of non-competitive assessment might accept'? Is there no truth in the normal curve theory of human performance and expectations? To which the answer, to the latter question about the normal curve theory, must be 'Quite possibly, but who wants to deal in generalisations when teachers have the opportunity to deal with individuals?' The process of learning is not to be reduced to the level of a discussion on the adequacy of measurement techniques. There are qualitative individual differences, and non-competitive assessment does not ignore the fact: teachers who use it are reaching positively to each individual's strengths, aptitudes and needs.

What functions might non-competitive assessment serve?

We believe there are five functions that non-competitive assessment serves.

Non-competitive assessment should

1. be *enabling*
2. promote individual *development*
3. be *co-operative* in nature
4. provide *assistance* to students
5. be useful in the *placement* of individuals.

We have deliberately chosen for these functions titles that are different from those used to describe competitive assessment in the first part of this paper, though we realise that the two modes overlap in some of their functions. It should also be noted that this categorisation is a proposal, not the result of reflection over a long history of assessment in this mode (as are the functions of competitive assessment). The headings are proposals of 'what should be' rather than reports of 'what happens'.

Enabling functions

Assessment should be enabling in various ways. It should enable

1. curriculum planners to make curricula student-centred rather than system-based
2. insertion into curricula of practice that fosters not only knowledge but also development of social and personal values, and a critical understanding or interpretation of knowledge
3. a process of learning and teaching to develop, in any school, that is responsive to the needs and abilities of both learner and teacher.

Development functions

It is a truism to say that a (or perhaps the) purpose of schooling is the development of each individual student to the greatest degree possible. Arguments will arise about what constitutes the most appropriate curriculum to direct and focus such development, but the point still stands. Assessment has the function of promoting and supporting this development.

Perhaps the key quality of non-competitive assessment in this regard is that it is not based on generalisations. It is integrated into the learning process, respecting that individuals start any sequence of instruction at different 'positions', that they have different values, have different rates or modes of acquisition of whatever is being taught, and that they finish the sequence of instruction just as 'different' from one another as they began (perhaps more so). One set of published guidelines for those who would write descriptive assessment statements makes the point:

Assessment must be integrated with the teaching/learning process. It should match the teaching/learning activities and foster further learning (VISE 1981:8).

There can be little doubt that a summative, competitive assessment system, such as one involving public examinations, simply cannot satisfy such a demand across the whole range of needs and possibilities available which it is incumbent on curriculum designers to consider. Concerning competitive, 'institutional' practices, Kleinig suggests:

The teacher passes on a pre-packaged and relatively fixed 'body of knowledge' to the student, and then determines how well that 'knowledge' has been assimilated. The assessment system thus leaves little room for a recognition of the social form and

content of that knowledge, or for an exposure and questioning of its presuppositions and values. Yet these are necessary if assessment is to have educational value. School knowledge functions as a more or less 'neutral' given, whose successful acquisition is unilaterally determinable. As such it constitutes (in many cases) a substantial constraint on the exercise or development of the learner's autonomy (Kleinig 1981:180).

Non-competitive assessment may be said to permit, perhaps even to encourage, the acquisition of such values and the acquisition of autonomy in the sense in which Kleinig seems to be using the term.

Assessment should also engender a sense of success in the individuals being assessed, whatever their cultural, social or educational background.

Competitive modes, of their nature, engender failure. Non-competitive modes can focus on the successes that learners have rather than their failures. Whatever the age of learners, all share a similar disposition – if they do well in their endeavours, or at least feel they have been successful, they may well be motivated to go on with learning. Should we not demand that our assessment system reflect this for all learners? This is not to discourage self-analysis or self-criticism as important values to the learner; the issue will remain primarily a matter of focus. There are more ways of telling people that they need to re-do or re-think something than saying 'Failed'.

Co-operative functions

If schooling is viewed as a social process involving in any one school (regardless of its 'philosophy') a large network of interpersonal relationships intended or structured to facilitate the attainment of educational goals, then the model of assessment used will contribute directly to such attainment. We believe the endeavours to achieve this end should be co-operative. Non-competitive assessment will promote rather than debilitate such co-operative endeavours, especially in learning situations where mixed ability grouping operates. Traditionally the focus or structure of the schooling process has centred on teacher-

student interactions, and these, as Kleinig suggests, are in turn dominated by the passive transfer of knowledge and values. However Apple goes further:

If our aim is a society in which all people are more equal in their opportunities to experience success and to exercise control over their own destinies, not a society in which the chasms between groups grow larger every day, then we must deal with these larger social issues. We must focus more vigorously – starting now – on the skills of democratic deliberation about such questions as social goals, the proper direction for schools to take, and what we should teach and why. We will never have a curriculum free of tensions and conflicts. And it would probably not be good if we did, since such conflicts demonstrate the vitality of democracy. We must learn to work creatively with conflicts, seeing them not as hindrances but as possibilities for cooperative improvement of education (Apple 1983).

But for a curriculum (and its assessment) to be truly co-operative, to reflect the maximum number of possibilities for learning inherent in the network available in any one school, it must encourage and recognise student-student interactions at all levels. Johnson, in a paper aptly titled 'Student-student interaction: the neglected variable in education', asserts:

Experiences with peers are not a superficial luxury to be enjoyed during lunch and after school. Constructive student-student relationships are probably an absolute necessity for maximal achievement, socialization, and healthy development (Johnson 1981).

He further claims:

The evidence suggests that the emphasis on teacher-student interaction reflects at least three major discrepancies between current educational practice and knowledge. The first discrepancy is that teacher-student interaction is emphasized and the possibility of student-student interaction is eliminated in many classrooms, while it is student-student interaction

that may be the more important determinant of educational success. The second discrepancy is that competition and individualistic learning dominate most American classrooms, while cooperative learning experiences appear to be far more effective in promoting desired educational outcomes. The third discrepancy is that, in many American classrooms, teachers generally avoid and suppress all conflict among students, yet spirited argumentation over academic issues is a powerful influence on student achievement, socialization, and development (Johnson 1981).

Non-competitive modes of assessment can account for the whole range of student-student interactions, and may go some way towards legitimating them. If a teacher's skills at fostering interpersonal relationships are highly developed, such assessment modes can use these relationships as part of their essential processes. Student assessment of other students' work will not be totally negative: sometimes there are insights available from these judgments which teachers would otherwise lack.

Assistance functions

Assessments can and should provide information to all learners which assists them in making decisions about their learning. Here we focus on the formative aspects of evaluating one's progress, as the word 'formative' is used in wider or large-scale evaluation contexts. This function is learner-centred and cannot reject the involvement of the student in the conduct of the assessment. For the assessment to be truly formative, the student must participate. Holding up the work of others as a criterion to be matched, or referring to the standards attained by others may well become 'competitive' elements of the process, even though there should never be any sense of such things being the focus of the review. The student needs to know what is to be learned, what is involved in achieving this learning, and how she or he is doing in the process of achieving it. The assessment can become part of the learning process, is not imposed by a teacher, and may well have spin-off in increased motivation on the part of the student. The assessment process can also be of use to the teacher, in supporting or refuting a judgment as to whether the course is suitable for the student.

This latter aspect is not, however, primarily what we mean by 'assistance' and perhaps comes closer to the final category.

Placement functions

We have some reservations about including placement amongst the functions of non-competitive assessment. This is not because it cannot serve a useful purpose in the making of such decisions – that is obviously possible. But we are here very close to the concept of 'selection' as it is currently used, the selection of a few from the many. Assessments using non-competitive modes or involving non-competitive attitudes on the part of the assessors are almost certainly more informative, in allowing a selector to make choices, than grades and marks. Such assessments say more about the individuals in the pool, say it more clearly and say it more accurately and validly. The irony of this situation is not lost on the present writers: that at the Great Competitive Moment in the lives of many Australian students – award of a post-secondary place – assessments arrived at through non-competitive means are to be preferred in theory (even if they are not, currently, preferred in practice) to those reached competitively. The unfortunate administrative difficulty of processing large numbers of applications for post-secondary places relying on descriptive assessments makes the use of competitive results the more common mode. However this should not be allowed to obscure the superiority of those reached by non-competitive means.

Burgess and Adams summarise the whole point:

We believe that what are increasingly required in the 14-16 range in secondary schools are not so much terminal measures of achievement to be used for selection purposes as kinds of assessment which provide teachers, parents, and pupils with guidance (Burgess and Adams 1980).

This leads us to another function we may reasonably term 'placement': using assessment to come to a decision about the appropriateness of particular future courses or career options. We do not need to demonstrate yet again in this matter the superiority of a participatory process yielding words rather than grade symbols or numbers.

Methods of assessment

While it is possible to be fairly precise about the functions of non-competitive assessment, it is not, given the nature of the assessment procedure itself, possible to propose a hard set of generalisations about how it should be conducted. Rather than a single test being applied to a class of 25 for the purposes of comparison (or a public examination to a cohort of 25 000), teachers may well find themselves conducting several or many different kinds of assessment within the one classroom. These will differ according to the different needs of different students or to the different stages various students are at, given the instructional program or the student work schedule, or whatever organisation is used to direct learning. The activities or behaviours central to the learning yield the bases for the assessment. The student and those others in his or her network (teachers, other students, other staff) become the assessors to varying extents. The statement of the assessment is determined by the audience for whom it is intended. These statements may be either oral (parent-teacher interviews) or written (assessment files, report cards, references). Above all they are not novel: teachers have been preparing them for years. What is new may well be the power and weight which they are now seen to have as replacements for, rather than random amplifications of, the numbers and the letters. Their value or power will depend directly on

1. the degree to which the statement is course-related
2. the extent to which the statement is negotiated, and the student participates in the negotiation
3. the precision with which aims and course details are described, to allow student progress or attainment to be evaluated by the readers
4. the degree to which the statement is audience-directed, to particular people or types of people
5. the variety of different assessment methods which led to the compilation of the overall statement
6. the sympathy and understanding with which the efforts of the student concerned were considered.

Guidelines exist to inform teachers and students about these matters (as in Thompson, 1978) and we will offer no further detail here.

Summary

In endorsing the view that assessment should be non-competitive, we have tried to say more than just 'assessment should not be competitive'. We have tried to offer a review of the functions of assessment (in both modes), to relate the functions to the processes and structures of schooling, to show something of the impact of each mode on learners and teachers, and to begin a discussion on the crucial way in which assessment represents the impact of society on its schools and schooling. We might then add the somewhat sad view that if society, and hence schooling, continues to be dominated by competitive actions and attitudes (and we may consider public examinations to be a tangible indicator of these), then non-competitive assessment is reduced to little more than a fashionable rhetoric, available to a lucky few.

Broadfoot (1982) has suggested that the development of public examinations, which have played such a central role in contemporary school systems, may be understood in terms of four themes: competence, content, competition, and control – competence attested to, content organised for, competition regulated by, and control provided through, public examinations. She elaborates this development by reference to the underlying ideologies.

At the heart of this development lay two developing ideologies. The first was that of 'meritocracy' – the selection of people for occupational roles not, as hitherto, on the traditional bases of birth and connections but on a new, more rational, basis of individual ability and effort. As people had accepted the old legitimisation of privilege, so they came to accept the new legitimisation of merit. Similarly they came to accept the second developing ideology – that it was appropriate for schooling, rather than some other social agency, to provide for this selection, by teaching particular competencies (later to become increasingly abstract and unrelated bodies of knowledge) and then 'measuring' who had and who had not reached a given standard. Central to the notion of 'merit' was that failure must be the

individual's, not the school's, fault and thus must necessarily be accepted by the individual (Broadfoot 1982).

The key issue in the acceptance of non-competitive assessment will be the degree to which it is superior to competitive forms according to the four aspects described above. Our argument has been that it is superior in the attestations of individual competence and in its relationship to the content of the learning. We have also suggested that at one point in the educational process – selection for post-secondary entry – it may, ironically, provide a better or fairer control of competition for places than assessments competitively arrived at. But such control is in the hands of teachers and students, not centralised public agencies, and until we accept that this autonomy is where it should be – a third ideology to add to Broadfoot's pair – then non-competitive assessment will find it hard to make its way.

Another obstacle to the ready acceptance of non-competitive modes across the whole of the secondary school has been the unwillingness of those who select for entry into post-secondary institutions to tackle the bureaucratic or logistic problems of coping with large quantities of descriptive assessment statements in arriving at their decisions. These problems have prolonged and will continue to prolong the operation of competitive grading practices. But as Kleinig remarks:

The purposes of certification and selection introduce into the learning and assessment situation psychological factors and other criteria (e.g. for ranking and grading purposes) which are not necessary to, and tend to be incompatible with, the learning process (particularly if it is to be educationally significant). What now needs to be done is for these two functions to be separated as much as possible. What this means is that accreditation should become the responsibility of employers, professional boards, etc., who can then conduct their own appraisals of potential employees, members, etc. Schooling certificates and basic degrees might then indicate only satisfactory or unsatisfactory participation in courses. There would not need to be any pass/fail quotas, grading or ranking. Not only would this improve the educational possibilities of schooling,

but it would also lead to more efficient selection, since employers, etc. would not have to put up with an assessment system pointed in a dozen directions (Kleinig 1981:191).

We also need to remind ourselves that no more than 35 to 40 per cent of secondary school students need to be affected by such logistic problems, and then only in the final, non-compulsory years of schooling. However non-competitive modes should not be neglected lower down the school simply to allow some students to practise at competitive examinations, held after the majority has left the school.

The practices of a competitive assessment policy have been promoted elsewhere on the grounds that, since competition is an element of life and since education is to establish in youth the necessary competencies of adult life, the competitiveness is acceptable. However if we take the view that learning is co-operative, active and personal, and if we are looking for logically concomitant curricula and assessment procedures, clearly we must move away from competitive assessment. Putting it another way, Beltz has urged:

Significant changes are necessary in schooling in order to meet more adequately the personal developmental needs of all students, and to equip them to operate in the world outside. The development within schools of more supportive and encouraging environments is essential. This should be brought about by examination of the structures and strategies that foster negative, authoritarian relationships. Encouragement of all students as individuals is essential, and practices which relegate many to lower status and lower expectations should be abandoned (Beltz 1980).

Such intentions are no more than pious hopes as long as competitive assessment remains a dominant mode in public or school policies.

References

- ANDERSON, J. J. 'Correlation between academic achievement and teaching success' in *Elementary school journal* no. 22 (1932) pp. 22-29.
- APPLE, M. W. 'Curriculum in the year 2000: tensions and possibilities' in *Phi Delta Kappan* (January 1983) p. 326.

- APPLE, M. W. 'Ideology, reproduction and educational reform' in *Comparative education review* (October 1978) p. 376.
- BELLANCA, J. A. and KIRSCHENBAUM, H. — 'An overview of grading alternatives' in S. B. Simon and J. A. Bellanca (eds) — *Degrading the grading myths: a primer of alternatives of grades and marks* — Washington: Association for Supervision and Curriculum Development, 1976.
- BELTZ, C. L. Educational responses to the changing need of youth. Paper presented at the 1980 ANZAAS Jubilee Congress, May 1980, p. 19.
- BLACHFORD, K. — *To 'B' or not to 'B' — Competitive assessment and some options: a research review* — Melbourne: Education Department of Victoria, Curriculum Services Unit, 1981.
- BLOOM, B. S. 'Affective outcomes of schooling learning' in *Phi Delta Kappan* — no. 59 (1977) pp. 193-198.
- BLOOM, B. S. *Stability and change in human characteristics* — New York: Wiley, 1964.
- BRIM, O. G. and others — *The use of standardized ability tests in American secondary schools* — New York: Sage, 1964.
- BROADFOOT, P. M. 'Alternatives to public examinations' in *Educational analysis* — vol. 4, no. 3, (1982) pp. 33-46.
- BURGESS, T. and ADAMS, E. — 'The present inadequacy' in Burgess and Adams (eds) — *Outcomes of education* — London: Macmillan, 1980 p. 5.
- CHAMBERLIN, C. and others — 'Did they succeed in college?' in *Adventure in American education* — New York: Harper and Row, 1942.
- CHANSKY, N. M. — 'The X-ray of the school mark' in *Educational forum* (March 1962) pp. 347-352 (Cited in Evans).
- CURWIN, R. L. — 'In conclusion: dispelling the grading myths' in Simon and Bellanca — *Degrading the grading myths* — *op. cit.*
- EVANS, F. B. — 'What research says about grading' in Simon and Bellanca — *Degrading the grading myths* — *op. cit.*
- GIROUX, H. A. — 'Hegemony, resistance and the paradox of educational reform' in *Interchange* — vol. 12, nos. 2-3 (1981).
- HANNAN, W. — 'Assessment: a workable, democratic method' in *English in Australia* — no. 59 (March 1982).
- HOYT, D. — 'College grades and adult accomplishment: a review of research' in *Educational record* — no. 47 (1966) pp. 70-75.
- JOHNSON, D. W. — 'Student-student interactions: the neglected variable in education' in *Educational researcher* — no. 10 (1981) p. 5.
- KAPPELL, F. R. — *From the world of college to the world of work* — New York: American Telephone and Telegraph Company, 1962.
- KIRSCHENBAUM, H. and others — *Wad-ja-get? The grading game in American education* — New York: A & W Pubs, 1971 (also available on Eric fiche ED 136436).
- KLEINIG, J. *Philosophical issues in education* — London: Croom Helm, 1981.
- PALLET, J. B. — Definition and predictions of success in the business world — Unpublished PhD dissertation, University of Iowa, 1965.
- PERSELY, G. W. — The effects of grading and previous academic achievement on subsequent academic achievement and student attitudes — Unpublished PhD dissertation, Bowling Green University, Ann Arbor, Michigan, 1974. *University microfilms* 78-34978.
- POLK, K. and SCHAFER, W. (eds) — *School and delinquency* — Englewood Cliffs, NJ: Prentice Hall, 1972.
- SIMON and BELLANCA (see Bellanca and Kirschenbaum).
- STARCH, D. and ELLIOTT, E. — 'Reliability of the grading of high school work in English' in *School review* — no. 20 (1912a) pp. 442-447.

- STARCH, D. and ELLIOTT, E. — 'Reliability of grading work in mathematics' in *School review* — no. 21 (1912b) pp. 245-295.
- STARCH, D. and ELLIOTT, E. — 'Reliability of grading work in history' in *School review* — no. 21 (1912c) pp. 676-681.
- TAWNEY, R. H. — *Equality* — London: George Allen and Unwin, 1951.
- THOMPSON, M. E. — Self-image: an important concept for study skills and professional development — Paper presented to Annual meeting of the North Central Reading Association, 1978 (also available on Eric fiche ED 163411).
- TIEG, E. W. — *Educational diagnosis* — Monterey, Ca: California Testing Bureau, 1952 (Bulletin no. 18) (Cited in Kirschenbaum and others).
- VICTORIAN INSTITUTE OF SECONDARY EDUCATION — *Discussion paper on guidelines for descriptive assessments* — Melbourne: VISE, 1981.
- WEST, I. H. T. and SLAMOWICZ, R. — 'The invalidity of the Higher School Certificate as a tertiary selection device' in *Vestes* — no. 2 (1976) pp. 8-11.
- WILKES, R. E. and NOBLE, T. E. — *The predictive validity of the 1966 Commonwealth secondary scholarship awards* — Hawthorn, Vic: Australian Council for Educational Research, 1972.
- YARBOROUGH, B. H. and JOHNSON, R. A. — 'How meaningful are marks?' in *Reading teacher* — vol. 33, no. 6 (1980) pp. 644-651.

Source: G. Withers & G. Cornish, 'Assessment in practice: Competitive or non-competitive?', Occasional Paper No. 11, VISE, Melbourne, 1984, pp. 2-16.

5

Year 12

Victorian Secondary Teachers Association

THE BACKGROUND

What to do about Year 12 has been a perennial question of Victorian education. VSTA policy on the subject goes back to 1965 when a motion of complete opposition to the Victorian Universities School Examination Board's (VUSEB) external examination was passed at an Annual General Meeting.

The intentions of this paper are to provide information about Year 12 which has been sought recently and persistently by a large number of schools, to illustrate the development of the Union's present policy and to assist with discussion and formation of new policy. A new policy is needed urgently to deal with changed circumstances, although the basic issues appear to remain unchanged.

These changed circumstances are several. Alternatives to HSC Group One subjects have proliferated as a means of completing Year 12. There has been a considerable growth in student numbers at that level, and hence there are more students who contest the relevance and value of the competitive academic curriculum. The job squeeze, coupled with increased aspirations for tertiary education has increased pressure in the credentialling race. And most significantly, there is now a VISE group reviewing arrangements for HSC.

Detailed reference will be made to these issues in this paper, but the one which has always been most pronounced is the balance in secondary education between the selection and sorting of students and the urgency of providing relevant and purposeful learning in less competitive situations.

The nature of the assessment process, the existence of the external exams, footholds gained in basing curriculum in schools — to mention just some of the conventional points of dispute — all stem from this larger problem; and lest we become insular about this, believing it to be a problem unique to Victoria, it should be pointed out that it is a matter of debate not just in other states of Australia but in every economically developed country in the world. Baldly stated, the question is this — who should have access, and on what basis, to the benefits of status and wealth that often follow from an extended education? Further, what effect does this have in the secondary sector on what we teach and how we teach it?

A BRIEF HISTORY OF PRESENT VSTA YEAR 12 POLICY

The policy decisions below have been summarised and grouped under relevant heads. They are those which stand at present. The consistency of attitude on these issues has played a substantial role in defining the ethos of the VSTA, a Union concerned not just with the industrial conditions of its members but with larger issues critical to the performance of their work.

Opposition to External examinations at Year 12

- AGM 1965 A motion of opposition to VUSEB external examination was passed.
AGM 1969 'That the competitive examination system be abolished.'
Passed.
This was reaffirmed at the 1977, 1979 and 1980 AGMs.

Assessment

- AGM 1974 'Oppose any assessment procedures which rank students.'
AGM 1979 Reaffirmation of the above.
AGM 1982 Assessment should be — goal based, continuous, diagnostic and participatory; 'no place for competitive assessment'.

School Based Curriculum Development

- AGM 1968 Curricula should be school based, i.e. devised, developed and taught.
AGM 1977 Reaffirmation of the above.
AGM 1978 Members urged to take maximum advantage of opportunities for school-based curriculum development under VISE.
AGM 1979 Confirmed total opposition to: 1. Curriculum uniformity, 2. Standardized testing, 3. Moderation of school assessments against external testing instruments.
AGM 1980 Reaffirmed school accountability for curriculum. Should be:
1. School-wide (c.f. intra-faculty) development.
2. Locally relevant with cognizance of a broader context.
3. Locally accountable.
4. Schools must have responsibility for assessment of students.
5. Broad, active participation in design.

Transition

- AGM 1979 VSTA urges schools to make use of diverse ways of gaining entry to further study and employment apart from externally prescribed and examined Year 12 subjects.

STC

- AGM 1977 Urging schools to join and support STC.
1978 Urging use of STC Principles, which are:
1979 1. The Group will require the participation of teachers and students and encourage the participation of parents in the development and assessment of each school's course as a whole and of individual subjects or units.
2. Content of particular courses will not be prescribed centrally.
3. The Group will not use or endorse any forms of assessment which grade student against student or school against school for purpose of selection of students to courses of tertiary study or employment or for any other purpose.

4. Entry to the STC Course will not be restricted on the basis on academic performance.
5. The Group will provide counselling aimed at ensuring that realistic and knowledgeable choices of work or tertiary study can be made by students.
6. Syllabuses may vary from school to school and from student to student but will be guided by the underlying values of relevance to the individual students' aspirations, social relevance (which requires recognition of the student's cultural origins and contemporary situations) and interdisciplinary relevance, which implies a continuing concern to relate disciplines one to another and to relate academically acquired knowledge to practical situations.

Standardization

Committee 1983: VSTA reiterates its long standing opposition to external prescription and assessment of HSC subjects and in particular the process known as 'standardization'.

It would be fair to say that these policies stem from the following broad considerations:

- objection to external control of the secondary curriculum, initially directly from Universities through an examination system, and later, following the abolition of VUSEB, *de facto* control exercised through entry requirements to some tertiary faculties;
- rejection of the suitability of a ranked, competitive assessment process especially when its use at Year 12 serves to establish the validity and importance of that sort of assessment throughout the whole secondary education;
- concern about gender, ethnic and social class biases in the groups which benefit from the HSC process.

It would be equally true to say that Union opposition to the arrangements for HSC prior to 1979 played a major role in causing those arrangements to be changed. The VSTA, however, was not responsible for the precise nature of those changes.

THE ADVENT OF VISE

(See Appendix 1. for a description of VISE structure)

The most significant of the new arrangements which the Victorian Institute of Secondary Education (VISE) instituted in 1979, was the setting up of a two-tiered subject structure. One tier consisted of subjects which had mostly existed prior to VISE and were called Group One. These subjects are externally examined and must of necessity have a core/option structure, with school assessment being responsible for up to 50 per cent of the 100 point scale. School assessments are however *moderated* against the score gained in the external examination. These scores are also *standardized*. (See below)

These subjects are administered by Subject Committees whose members are appointed by the Curriculum and Assessment Committee (C&A) (See Appendix 1.) usually on advice from the incumbent members. The relationship between these committees and the central VISE structure is variable and sometimes stormy.

The second tier consisted of Group Two courses — single units, subjects or Approved Study Structures (ASS), which could be offered to VISE for accreditation by any interested person or group and which are available for any school to teach.

The accreditation process is formally the same as that for Group One subjects, although in practice it is often more stringent. These courses do not have to conform to the core/option structural requirements of Group One subjects and are entirely school-assessed. In a minority of these subjects there is a moderation procedure through teachers' consultative meetings.

An error often made is to refer to 'HSC subjects' to mean Group One subjects alone, distinguishing them from 'Group Two subjects' or even more specifically 'STC'. *Group Two studies are one means of participating in the HSC program, not something quite separate.* 'STC' which stands for 'Schools Year 12 and Tertiary Certificate' is also known as 'Approved Study Structure V', and is a VISE HSC Course, formally certified in a way no less authentic than any other VISE subject, unit or ASS.

A further common mistake heard constantly at school meetings held to discuss Year 12 offerings, is that Group 2 courses are not intended by VISE as tertiary entrance prerequisites. *This is not true at all.* All documentation and procedure suggests that *each VISE accredited course should be accorded equal status.*

People are also often puzzled about *Tertiary Orientation Program (TOP)* courses. These do not come under the aegis of VISE, but of TAFE and are administered by TAFE teachers grouped regionally, in collaboration with a Technical Institute (e.g. RMIT, Footscray) or a College of Advanced Education (e.g. Bendigo). It is widely known that student numbers in these programs have grown quickly and substantially. (In 1982 total enrolments in HSC English were 21,396, total TOP enrolments were 8,652). It is less well known that VUAC discounts TOP score aggregates in the allotment of tertiary places.

Group Two courses

It was commonly assumed early in the history of VISE that Group Two subjects would wither on the vine for want of takers. This has certainly not occurred. The number of new Group Two offerings exploded in 1980-82 while tailing off more recently (see Appendix 2. for a complete list). The growth in the number of both schools and students involved with Group Two courses is dramatic. As the figures below demonstrate, a large number of schools now offer one or more Group Two courses.

FIGURE ONE

School Type	NUMBER OF SCHOOLS OFFERING AT LEAST ONE GROUP 2 COURSE		
	1981	1982	1983
Catholic Secondary	14	36	50
TAFE Colleges, Evening High Classes	1	3	10
Govt High-Technical	0	1	6
Non Catholic, Non Govt Secondary	16	24	30
Govt Secondary	46	77	116
Govt Technical	2	3	3
CAE, Community Centres Learning Exchanges	0	2	3
TOTAL	79	146	218

FIGURE TWO

**YEAR 12 CURRICULUM 1983
BY SCHOOL TYPE AND COURSES OFFERED**

School Type	Number of Schools			Total
	Group 1 only	Group 1 and 2	Group 2 only	
Catholic Secondary	33	50	0	83
TAFE Colleges, Evening High Classes	35	9	1	45
Govt. High-Technical	8	6	0	14
Non Catholic, Non Govt. Secondary	35	30	0	65
Govt. Secondary	174	109	7	290
Govt. Technical	3	2	1	6
CAE, Community Centres Learning Exchanges	9	3	0	12
TOTAL	299	209	9	517
(No. Students)	(23,675)	(3,523)	(1,500)	(28,698)

The number of student participants also shows rapid growth, more than doubling in each year for which figures are available.

Approximate student enrolment in all Group Two courses.	1981	1982	1983
(unit, subject & ASS)	1,300	3,700	8,850

In 1984 it is estimated that more than a quarter of all HSC students will be taking at least one Group Two unit subject, or an Approved Study Structure.

Schools have taken on Group Two for various reasons but the following are the most common:

- to improve their upper school retention rates;
- to broaden their Year 12 curriculum and thus cater more satisfactorily for students' needs;
- to retain control of assessment procedures; and (the one which is of most concern)
- to stream off students who are perceived to be less academically able into an 'easy' course, thus keeping up pass rates.

It is incorrect to talk about Group Two courses as though they were all the same. 'Maths at Work' and 'Commercial Maths', for example, are nominally similar but in fact are very different in the way the courses are shaped, in their assessment procedures and in the sort of educational experiences offered to students.

A majority of Group Two courses are not dissimilar from Group One courses. To offer an example again, there are a number of foreign language studies in Group Two because they were recently devised and for no other reason.

A majority of Group Two courses have examinations and graded assessments and some are more rigid in their requirements than Group One subjects, all of which have been leavened to some extent by the core/option structure.

There are distinct internal anomalies in some Group Two subjects particularly in the consistency of the stated aims and the nature of the assessment. Perhaps the most glaring are to be found in the units 'Personhood' and 'Christian Prayer' both taken by a substantial number of students, which are assessed on an A-F scale.

It is pleasing to note that amongst the largest and fastest growing subjects are several which are genuinely innovative and flexible with assessment arrangements which are in accord with VSTA policy. These include Creative Arts, English B, Maths at Work and STC. (See Appendix 3. for precise rates of growth).

While becoming increasingly rare, one still finds Principals or teachers puffing out their chests with pride that 'they have stayed a Group One school'. *At present the situation is such that the very wide range of offerings for HSC should be seen as a catalogue to draw from, to best suit the needs of the students within a school according to the resources available at that school.* Argument, tinged with bitterness, about the propriety of Group Two offerings is from another time and another set of circumstances.

But there is a natural concern about students' futures which must be dealt with.

WHERE DOES GROUP TWO GET YOU?

This is the most common question asked about Year 12, and the most difficult to answer with any certainty.

It is important to say that one should have some sympathy with tertiary institutions faced by a welter of alternatives, and who find it natural to seek out the most clerically simple course of action for selection, despite considerable qualms about its validity and relevance.

Most recent information from VUAC (February 1984) suggests that there are 39,000 applicants for 19,000 tertiary places, of whom 25,500 are HSC/TOP candidates and the rest are made up of mature age, interstate or overseas applicants (the last having risen by 16 per cent from last year), and transfers from other courses. This indicates a very tight squeeze, with a lot of suitably qualified applicants disappointed. Yet recent figures suggest that students who have taken one or more Group Two subjects *do gain entry to tertiary education*, and in substantial numbers. (See Appendix 4. VISE Survey of Placements to Post-Secondary Institutions), but it is a much more arduous process.

The Guide contained in Appendix 5. indicates some of the difficulties associated with offering accurate counselling to students. There are some further matters which should be borne in mind.

1. There are a great many post-secondary courses on offer. There are ways into first choice courses even if they are not gained initially, including enrolling elsewhere and transferring after completion of a year or a number of years of the course.
2. The Anderson score (an aggregate of four best standardized HSC scores combined with ten percent of the fifth) is by no means the only mode of entry. Of 240 courses controlled by VUAC (and there are literally hundreds of others) 63 have

selection procedures which are not solely based on scores, and many of the others have quota requirements for, for example, mature age students. In 1982-83 10 of the 240 admitted all eligible applicants.

Although it sometimes sounds as if there is no tertiary life apart from doing Medicine at Melbourne, that is not necessarily the case.

THE PRESENT HSC REVIEW AND FUTURE DIRECTIONS

Around the beginning of 1983 a group was formed at the instigation of the VISE Curriculum and Assessment Committee to deal with matters — some longstanding, some which have come out in the wash since the Year 12 changes and some which were quite novel — which raised questions about VISE procedures. There were three large issues — the standardization of marks, the grouping of subjects and the recommendations of the Report concerning English as a Compulsory Subject — and several other minor ones.

The group, made up of representatives from each of the major subcommittees of C & A and including the Chairman of VISE, Bernard Rechter and the Head of the Curriculum and Assessment Branch, Win McDonell, has now completed their second report which will be forwarded for consideration by the C & A Committee. But before going over its recommendations, a description of the issues is in order. Necessarily, what follows will be a simplification, but it is important that these issues be widely understood and considered.

The making of a mark

It is easy to assume that kids sit their Group One Subject papers, the examiners mark them out of, say, seventy against an enduring, clear standard, the mark out of, say, thirty that teachers have given their students is added, and then the percentile is forwarded to VUAC as part of an Anderson score.

This is not what happens at all. At least four other processes intervene between the initiation and the outcome.

The first is *moderation* of the school-assessment against the exam score. If there is much deviation between the two the school assessment will be discounted. (See Appendix 6. 'Moderation'.) The clear assumption is that *teachers' assessments are less valid than those produced by the examination*.

Then after a number of fiddling operations the marks will be *standardized* against the students' performances in each of their other subjects. This process has been written about at length in the past (See Appendix 7. 'Standardization') and has engendered very considerable debate. The assumption is that there are 'hard' subjects and 'easy' subjects which can be established by this process. A by product of this has been the discovery that Maths/Science students perform fairly consistently across their subjects.

Then the 'marks' which have already come a long way from their original form are *scaled* as part of the standardization process. The computer will have allotted a percentile of students to each letter grade as the basis of the standardization.

The students who got 158 out of 200 for English last year were assigned 100 percent. The students who got 79 or more out of 200 for Applied Maths got a D,

with just below eighty three percent of all Applied Maths students getting a D or above. Legal Studies in 1983 was a very tough subject with poor students. It must have been, because only 62.4 percent of students (with 106 marks out of 200) got a Grade D or above.

Lastly, there is a human factor called *The Standardization Meeting* where all the information and advice provided by the computer is juggled by members of the VISE Assessment Procedures Committee.

If a case can be made that this process ensures the maintenance of educational standards of performance, then it has yet to be discovered by this writer.

The Labor Government has a policy demanding the scrapping of standardization which has not yet been enacted. A discussion paper on "Standardization" was issued by the VISE some time ago, but no recommendations were forthcoming from the responses. This was one issue assigned to the Review Committee to work on.

Subject grouping

As has been mentioned, VISE grouped subjects initially on the basis of course structure and assessment procedures, with a bottom line being a maintenance of "old" courses in Group One with a possibility of the introduction of new courses into the Group Two mode.

A number of courses submitted for accreditation since the inception of this system have sought Group One status, but there has been no clear basis for making this decision. Non-VUSEB courses which were included in Group One very early in VISE history were Computer Science, Physical Education and Secretarial Studies. Since then only two subjects, Serbo-Croatian and Textiles, have been admitted to Group One — the first as a result of effective lobbying, the second as a subject more inclined to be taken by girls.

The attachment of status to course grouping has naturally raised questions of the justice of the grouping, but there is also the matter of rationality. Why is Lithuanian, which in 1983 had six candidates, in Group One, whereas Slovenian, which in 1983 had the same number of candidates, is in Group Two?

There was also the question of cost and administrative efficiency. Of the fifty four Group One subjects offered in 1983, twenty nine had fewer than 500 candidates, seventeen fewer than 100, and ten fewer than fifty. It is very costly per capita to mount external examinations for subjects with fewer than 100 students.

Further, if *all* subjects have to undergo the same sort of accreditation process why should they then be separated in perceived status by being placed in different groups?

PROPOSED CHANGES

A working party to review HSC has now presented its report to the Curriculum and Assessment Committee of VISE which passed it on substantially unchanged. We understand that VISE Council has also accepted the recommendations and intends to circulate them, with explanation, for extensive public discussion. The earliest year for operation of the proposed changes would be 1987.

The principles on which the proposed changes will be based stress that Year 12 courses should be primarily concerned with the promotion of learning, and that the demands made on HSC as a credential should be subordinate to that. This principle is balanced by a reaffirmation of the major role VISE has always had: to try to achieve, as far as possible, comparability of content and assessment in Year 12 courses.

To reconcile these principles, the proposals before VISE Council advocate a shift from norm-referenced assessment towards criteria-based or work-required methods of assessment. The present categories of Group 1 and Group 2 would disappear, and all subjects would be reported on with either a five-point scale (A-E) or a two point scale (S/N). 'Small' subjects (ones with less than ten per cent of the total candidature) would be totally school-assessed. 'Big' subjects (with more than ten per cent of the total candidature) would be assessed half externally and half on the basis of school assessments. School assessments and external assessments would be equally weighted. The method suggested for weighting different assessments from schools and external examiners can be summarised in the table:

		External Assessment				
		A	B	C	D	E
Internal Assessment	A	A	B	B	C	D
	B	B	B	C	C	D
	C	B	C	C	D	D
	D	C	C	D	D	E
	E	D	D	D	E	E

Thus, if the internal assessment is C and the external D, the final assessment is D. An internal C and an external A would result in a B. And so on.

All subjects with an external component would be assessed on a five-point scale.

To ensure comparability the proposals require a form of moderation with each accredited course. The preferred form is consensus moderation or consultative meetings of teachers taking the subject. Standardization as a form of achieving comparability is to be dropped.

The proposals envisage that Approved Study Structures and Single Units will continue, although changes may be necessary to some to make them compatible with the new system.

ENGLISH AS A COMPULSORY SUBJECT

A Report on English as a Compulsory Subject was presented to the Curriculum and Assessment Committee in May last year and referred to the HSC Review Group. The Group and Committee accepted the substance of the recommendations in the Report but modified them to fit the pattern for HSC proposed by the Review Group. Basically, satisfactory completion of a Year 12 course will require a satisfactory grade in English. There will be no provision for compensation without a Satisfactory Grade in English. A new English course is to be designed as soon as possible by a Course Design Group. It is to have a single structure which caters for the fullest possible spectrum of English students and teaching strategies. Assessment requirements must stipulate explicitly what work is required for satisfactory completion and the report should amplify the basic S/N assessment.

WHAT DOES THE VSTA DO NOW?

That, of course, is a question for its members to decide. The issues, which may seem to be increasingly technical and complex, still relate strongly to the main concerns of extant policy.

The largest question is still — should Year 12 be a year of general, common education for those students who wish to undertake it, unencumbered by the stringent controls urged on Year 12 by those who wish to use it and its products as a mechanism for selection?

The second question is a more specific spin off from that — many rigid paths, a number of which lead nowhere, or one flexible path? VSTA policy for many years has been totally opposed to streaming. The situation we find ourselves in, with Group One and Group Two, is presently being used for streaming in its most potent form, and tying both teachers and students up in the welter of alternatives, not to mention providing no clear incentive for tertiary institutions to see all Year 12 offerings as being of equal value.

The government's designs to keep kids at school rather than on the dole have been among the factors which have exacerbated this difficulty. There is no sign yet of any local legislative pressure being brought to bear to institute more thoughtful and democratic means of entry to tertiary study, although Senator Ryan has instructed the Tertiary Education Commission,

"... to bring forward recommendations proposing ways of achieving rapid, substantial and sustained reductions in the mismatch between the composition of society and the social composition of tertiary institutions, individual faculties and the tertiary sector as a whole. Recommendations should pay attention to the present nature of requirements for entry, *especially* the use of exam-based credentials and possible alternatives to them, the provision of bridging, remedial or introductory courses or the provision of facilities such as counselling services, support groups and transport arrangements to meet the special needs of disadvantaged groups."

It is obvious, as it has always been, that the present Group One subjects do not meet the educational needs of many students, in fact probably not of the vast majority. It is necessary to make a fundamental discrimination between educational needs and instrumental needs. Instrumental needs, in this context, tend to be the overriding factor. But there is no explicit equation which guarantees Group One = Tertiary Education = Wealth and Status. Even if there were, it could be that those final goals do not represent what is good about learning and developing students' potential, intellectually and in other ways. Members of this Union have an explicit professional responsibility to attend to those educational needs.

With these things in mind the following draft policy on Year 12 has been prepared for discussion.

DRAFT POLICY — YEAR 12

1. The Purpose of Year 12 Education

- 1.1 The central purpose of Year 12 courses must be to educate all students enrolled in them rather than to provide procedures for selection. This principle should be pre-eminent in any policy deliberations.
- 1.2 The range of courses should emphasize general rather than specifically vocational education.

The first clause of this section appears to be beyond dispute, and reflects VSTA policy as it has existed over many years. The second is perhaps more arguable. It is claimed by some that a 'liberal' or general education is inappropriate to some or many older adolescents who want and need to get their teeth into more concrete activities. The argument does depend, however, on the nature of the 'general' education. Few of the subjects conventionally considered to be a part of general education would claim to be aloof from practical activity as part of the learning program, and many would be considerably enriched by such a development. Similarly, intellectual development should not be seen as the preserve of the few.

Further rewording and discussion may produce a more satisfactory version however.

2. Opposition to current practices

The VSTA declares its opposition to

- 2.1 practices which contribute to the grouping of subjects in a hierarchy of prestige;
- 2.2 norm-referenced assessment procedures;
- 2.3 all forms of statistical moderation;
- 2.4 the process known as 'standardization';
- 2.5 the process of common scaling;
- 2.6 the current practice of score aggregation, used for the selection of entrants to tertiary institutions and other similar purposes.

This section collects and reiterates former policy statements. (2.1) is the product of experience which finds students and teachers who undertake Group Two courses still at a number of unjustifiable disadvantages. It is phrased generally because, if subject grouping persists, what the clause refers to can be expected to reappear in various and possibly unanticipated guises.

The next four clauses refer to the procedures which intercede between the initial assessment (in a 'raw' state) and the finally established score of the majority of students. The Union's point of particular opposition has been and remains the invalidity of the assumptions on which these procedures rest.

The last clause is seen by many to be the most important.

Score Aggregation reinforces the concept of selection: based on a notion of 'general ability', paradoxically producing hierarchies of subjects and courses — the 'true villain' as it has been described. We know that it does not predict tertiary success very accurately, in fact it does so less well than teachers' opinions.

The current 'Anderson Score' procedures based on the aggregation of scores on a hundred point scale produces bizarre anomalies. Why, for example, should students

with a score of say 296 be prohibited from entry to a course which has a cut-off point of 297, particularly when consideration is given to the process by which that score is created?

3. Course Design and Accreditation

- 3.1 All courses should be centrally accredited according to a set of common guidelines concerning definition of the students for whom the course is intended, course design and assessment procedure which should include the following changes:
 - 3.2 There should be a regularization of the format of all VISE course descriptions which would clarify:
 - the intention of the course
 - description of the method prescribed to achieve those intentions
 - an assessment procedure which is consistent with the principles of established VSTA Assessment Policy (see below).
 - 3.3 Accreditation procedure should ensure precise consistency between those three elements.
 - 3.4 All groups offering subjects for accreditation should be able to offer evidence of substantial input from practising secondary teachers.

The first three clauses are directed towards the improvement of course design through the accreditation process in three ways — regularization, internal consistency and accessibility.

At present, course descriptions can in no sense be described as 'regularized'; they are extremely various. For example, a senior member of VISE staff has identified no less than sixteen different assessment patterns for Group One subjects which nominally have an identical assessment procedure.

Further, while the design layout of courses has been undertaken with care the textual arrangement of Course Descriptions varies considerably.

The lack of internal consistency, the most serious matter, is a feature of a large number of both Group One and Group Two courses, predominantly between the stated aims and intentions of the course (which are often warm, encouraging and open) and the assessment procedure (which is not). Of those two items it will be the assessment which will provide the focus for the conduct of the course. There are also manifold inconsistencies with both those items and the way the 'middle' of the course relates to them. Apart from being sloppy and inconsistent there is a real danger of consumers being taken in by inaccurate labelling. More stringent requirements should exist.

Both improved regularization and internal consistency will help to make course designs more accessible. With the very large number of course possibilities on offer, access for everyone concerned with course choice should be seen as being of critical importance.

Regarding the provision concerning assessment — if we have a policy on assessment then we should make every effort to implement that in larger and perhaps more significant arenas of assessment.

In the past, the problem has been the recognized necessity for some form of selection procedure to take place at Year 12. Several formulations have been offered

to deal with this in constructive and thoughtful ways. Further work has confirmed the profound difficulties associated with this task.

However we believe that the propositions regarding 'Tertiary Entrance', (6) of this paper, make following through our assessment policy realistic and attractive. It is suggested that this section should be read in concert with the policy on Assessment (below), which was passed almost unanimously at the 1982 AGM, and the tertiary entrance section.

The last clause has been included as a matter of principle — practising teachers should have some input into course design. They do not necessarily at present.

4. Moderation

- 4.1 There should be provision made for moderation of all Year 12 courses in a consultative, consensus mode. Any reviews of assessment, or appeals against assessment should be conducted and controlled by the moderators who shall be practising secondary teachers.**

This section is based on three considerations. The first is that no teacher should be left to carry the responsibility for assessment alone. Confirmation of judgements and shared responsibility are crucial to the success of school-based assessment.

Secondly, there is very considerable evidence of the value of this form of discussion as a model for genuine school-based curriculum development, hence the reference to 'consultative'.

Thirdly, where such positive aspects exist it is useful to be able to accede to requests from tertiary institutions for moderated assessments.

5. Certification

- 5.1 There should be a single certificate covering all Year 12 courses which:**
 — is based on the concept of *completion* of the goals referred to in present assessment policy;
 — would register 'completion' or 'non-completion', and include a brief descriptive assessment or profile for each subject (unit or course) undertaken.
 — may be cumulative (i.e. there should be no specified time limit);
- 5.2 The VSTA would welcome the inclusion of Tertiary Orientation Programs and any other Year 12 courses within the single certification, and hence accreditation, procedure.**

This is a logically and practically necessary extension of Section 3 regarding Course Design.

The second clause is again designed to introduce further order to current Year 12 arrangements. It also confirms a recommendation of the Ministerial Working Party into Technical Education, and would remove the present disadvantage suffered by TOP students.

6. Tertiary Entrance

- 6.1 Consistent with principles of participation, equity and access, selection into specific faculties and courses should be undertaken by tertiary institutions following a year of general education administered by and taught within those institutions, relevant to prospective further studies and open to all who are suitably qualified.**

- 6.2 Qualification for entrance to that first year of tertiary general education should be on the basis of:**
- completion of course work;
 - completion of studies in prerequisite areas as defined by a body controlled jointly by representatives drawn from tertiary and secondary education (including substantial representation from practising secondary teachers) and other relevant community groups.
 - descriptive assessments or profiles;
- 6.3 Where applicants remain in excess of quotas, selection should be made by the following means:**
- rationalization of courses and subjects offered at tertiary institutions, with a view to limiting competition between institutions in the same kind of courses;
 - expanding places to meet demand from applicants;
 - establishing and filling group quotas (e.g. gender, ethnic group);
 - use of interviews and other instruments (such as folios of work, performances, evidence of experience) designed to gauge personal suitability.

This could be the most fertile proposal contained in the draft. It introduces the notion of a Year '13' as the basis for continuing tertiary education. It is not a novel perspective — the possibility of this was drawn to our attention by various comments made by or attributed to representatives of tertiary institutions.

At a time when the credentials race is having an increased impact on secondary education, retention rates are increasing and as a consequence far more pressure than ever before is being placed on Year 12, it seems appropriate to push back that selection procedure a further year if only to allow the proper implementation of another VSTA policy which states 'All students have the right to complete six years of secondary education, which includes the right to complete a Year 12 course'.

At present this is generally occurring, although a number of schools remain which have extensive culling operations at the end of both Year 10 and 11. But it is the manner in which it is occurring which is problematic. The spectre of the 'terminal course' hangs over a number of the alternatives on offer. This encourages schools to abdicate a certain amount of their responsibility for the education of *all* their students, and assign fewer resources and vastly less prestige to 'alternative' course streams. It encourages considerable division between schools operating in the same area, and places a completely unacceptable obligation on those involved in secondary education — that is to focus their efforts on the success of the few rather than the education of the many.

Tertiary institutions have consistently claimed proprietary right consistently over tertiary entrance, and this seems a means of satisfying this demand. It would offer them responsibility for their own judgements (and mistakes) in continuing selection, and (according to research on the subject) offers better predictive capabilities for assessing potential student success.

From the point of view of government secondary schools it is in useful accord with research which indicates that, while students from those schools find it harder to gain tertiary entrance, once having entered they do better through the following years of the course than students from non-government schools.

It enables those involved with secondary schools to control secondary education in an unprecedented fashion, alleviating some of the growing pressure on the curriculum. It ought also to make it easier for the old divisional boundaries to disappear, at least more quickly than they otherwise may have.

We do not believe that tertiary institutions would object to having their numbers expanded or to making the relatively minor changes necessary to establish a year of general education related to their faculty structures. In fact many faculties already treat their first year in this fashion. That is not to mention the educational advantages to students.

A very considerable number of applicants to first year courses are 'transfer students', that is students who are presently undertaking a different tertiary course. Not only does this suggest the wisdom of the institution of a general year, but it also means that this group of students would be deleted from first year entry, reducing numbers in that group quite dramatically.

Cost is a factor, probably the factor, and yet what is suggested above would be a great deal less costly than recent TAFE initiatives (with which they would be entirely consonant) and which at least establish a precedent for spending on that sort of educational program. The suggestion is also entirely in conformity with Senator Ryan's recently issued Guidelines on Tertiary Entrance with relation to access and participation, if not equity.

Equity is the concern of (6.2) and (6.3). These suggestions or ones like them have been mooted for some time, although they vary substantially from former propositions regarding Open Entry. To that extent we believe that they are 'known of', will not come as a shock and are fairly self-explanatory. They are also realistic and in conformity with (6.1).

The draft is commended to you for discussion and constructive amendment.

VSTA Policy on Assessment

Preamble

As indicated in 1980 policy 'control of assessment is tantamount to control of the curriculum'. This position recognized the power of external examination to control the practices of school curriculum. Thus it is important to establish that the form of assessment must depend on and flow from the nature of the curriculum. Assessment serves the purposes of curriculum.

The 1974 AGM established a basic framework for student assessment when it stated:

The VSTA opposes any assessment procedures which rank students, but encourages individual-centred methods of teaching and reporting.

School experience has led to formulation of curriculum methods that are sympathetic and consonant with such procedures.

Policy

School courses should be constructed around goals which all students can achieve. Within these goals, decisions about student learning activities should be negotiated between students, teachers and parents with the participation of the wider school community where appropriate.

The purpose of student assessment is to provide information on the progress that student is making towards achieving these goals.

Assessment procedures should be continuous, diagnostic and participatory. Summative assessments based on the continuous assessments should be provided at appropriate times during the year and at the end of the year.

Such assessments should at all times be available to the individual students and their parents.

Continuous assessment means that work throughout the course should be taken into account in making decisions for the purpose of assessment. As well students should have a clear idea of how they are performing in relationship to established goals — at any given point in the course. Any summative assessment should reflect the continuous assessments.

Diagnostic refers to analysis of strengths and weaknesses of student performance, in relation to established goals and directions suggested for improvement.

Participatory refers to the requirement that assessment processes are developed by mutual agreement, are properly understood by all concerned, and are carried out in fully collaborative manner.

To elaborate these assessment procedures the following principles are adopted.
Assessment

(i) The VSTA supports the notion of learning as an essentially co-operative endeavour. There is no place for competitive assessment in this model of education.

(ii) Assessments of students should be written and/or verbal accounts of work completed, learning processes, skills and knowledge covered, course objectives met and not met. These descriptive accounts, drawn up by teachers and students would state clearly what the student has done and what the student should proceed to do.

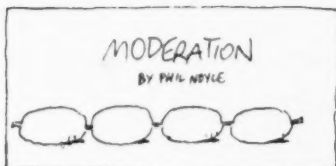
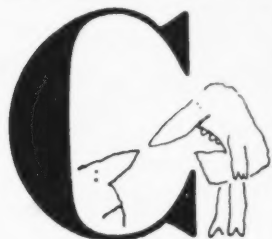
(iii) Assessment statements should not be based on any arbitrary year-level standards but should be based on the student's attainment of appropriate goals.

(iv) Summative assessments should take into account all aspects of a student's development, including a student's self-assessment.

Appendix 6

Curriculum

From *The Secondary Teacher* No. 8, 1980



'MODERATION is one of the major conflicts between political expectations and educational philosophy.*

It is a conflict because the political world expects marks at HSC to be 'comparable' whilst educationists are concerned with development of the individual.

Whereas at all years below Year 12 people don't seem to worry whether an assessment at one school means the same as at another, at HSC they do. And because they do, any school-based assessment in a particular subject will have to be 'moderated' in some form or other. That's the price to be paid for the school gaining some measure of curriculum autonomy; some method for getting comparability of different assessments of student performance statewide in a particular subject has to be imposed.

From 1981 on, every Group 1 HSC subject will have a minimum of 30 per cent and up to 50 per cent school-based assessment. It's to be split into two categories:

Teacher assessment of core (up to 20%)

About half of next year's Group 1 subjects will have a 10 per cent or 20 per cent weighting for teacher assessment on the content covered by the exam.

* Jay D. Scribner and Richard M. Englert, *The Politics of Education*, p. 22, 1977. (American) National Society for Study of Education.

Teacher assessment of options (30%)

Many skills and abilities can't be adequately assessed by an instrument as remote as the external exam. Whereas in the past they have simply not been acknowledged, it's argued that school-based assessment of options will strengthen and broaden the total assessment.

Finding some way of giving a mark the same meaning from Moonie Ponds or Munungatung, whether given by an easy marker or a standards stickler, is the task of moderation.

Broadly, there are two ways of going about it:

1. **Statistical moderation** is the term used for applying some external instrument (for example IQ test, aptitude test or exam) to the teacher assessment so that it will conform to the marks that class as a group got in relation to the rest of the state.
2. **Consensus moderation** describes any one of a variety of methods where teachers attempt to compare assessments.

Statistical Moderation

In Queensland and the ACT, where no external exam exists, the Australian Scholastic Achievement Test (ASAT) is used. As a moderating instrument, it is too suspect for anyone in Victoria to suggest we adopt it. In any case, we have something ready-made and possibly better-suited — the HSC exam.

WISE policy is to statistically moderate teacher assessment against the exam in the following cases:

(a) Wherever a subject includes a teacher assessment of core component. It can be argued that as the same content and objectives are being tested but from a different vantage point, it makes sense to regularise the marks against the exam.

(b) On the option section unless a request by the subject committee to use consensus moderation is approved by VISE Council. In general this can be defended only on economic, political and administrative grounds as options are entirely school assessed, often school devised and constructed to test different skills and abilities from the exam.

What Statistical Moderation does to the teacher's assessment

Marks are altered in two ways:

1. The mean is shifted so that the average of the teacher's assessments for the class is the same as that for their external exam results.

2. They are moved closer together or further apart to conform to the spread that that group got in the exam.

Whilst the marks are changed however, the rank order is unaltered.

The purpose of this mark processing is to render the teacher assessment amenable to addition to the exam score prior to the sum of the two being 'standardised' (the process that makes scores between subjects 'comparable'). By this stage, both teacher's and examiner's marks are little more than a gleam in the diode of the computer.

Perhaps the most damaging aspect of this mechanistic approach to measurement is that at no stage is information gained by or about teachers' marking methods. Teachers (and examiners, for that matter) are locked out of the process. They get no feedback on criteria or techniques. School tests badly constructed will not be picked up, diagnosed or remedied. Teachers will not have driven home to them that their assessments of students are inextricably bound up with the assessment procedures they employ.



What advice teachers will get out of statistically moderated assessments was recently decided by VISE as follows:

That schools and teachers should receive reports on the 'accuracy' of their assessments in the following form:

'The mean percentage score of your class on the external examination was X. The mean percentage score of your class on the school assessment was Y. This suggests that compared with the external assessment your assessment was too hard/about right/too lenient.'

The standard deviation of the scores for your class on the external examination was A. The standard deviation of the scores for your class on the school assessment was B. This suggests that compared with the external assessment your assessments were too bunched/about right/too spread out.'

Subject..... Date.....'

What isn't made clear is that it doesn't matter a fig if next year the same teacher's marks are any different: the machine will sort it out again.

Going back one step: Is it legitimate to add moderated teacher assessment to the external exam score? Are they measuring the same things? The answer depends largely on whether the marks correlate, that is, if the teacher's ranking compares well with the exam rank order.

(Statistical aside: A correlation factor of +1 means the exam and the teacher assessment produces exactly the same order of results. Similarly, -1 means a perfect reversal, and 0 means that tossing a coin would have done as well.)

A strong relationship (high correlation) we could expect, will be achieved by testing the same things as the external exam. Conversely, if different skills and abilities are tested, especially those things that the exam cannot test, a negative correlation could result.



Does VISE expect high or low correlation?

A limit is set, below which the teacher assessment is simply thrown out and the exam takes up the full mark range. Currently, that limit is a correlation factor of +0.1.

This means that, in terms of prediction of teacher assessment by the exam, only 10 per cent of the teacher's marks can be explained in terms of the examiners' marks. That is, 90 per cent can't be.

This exacting limit (slightly better than random) is not always met. The policy in such circumstances is simply to wipe the teacher's assessment and scale up the exam to the full amount.

There are a number of casualties produced in this inexorable quest to squeeze a year's work into a single number. Measurement theory is one of them, teacher involvement in technical issues of assessment, and as a consequence, curriculum development, another.

We have, in a sense, a *Catch 22* situation: if the marks don't correlate well, it's not legitimate to add them up; if they do correlate well there's no point in having teacher assessment.

The measurement experts scramble out of that dilemma in the following way:

"There are two reasons for adding unlike things:

1. When they measure the same skills from different vantage points the sum can give a more reliable result.
2. When they measure different skills a composite can give a better description of overall performance."

The argument reminds one of the blurb used to promote the health food which was "ideal for athletes, growing children and adults".

The real question raised by the fact that the teacher assessment is overridden by the external exam is: Why bother with the school-based options at all? For options to be anything more than a device to give teachers the illusion of participation, the use of statistical moderation is suspect, working as it does on the assumption that the external exam is correct.

Consensus Moderation

There are many ways of approaching comparability without assuming the external exam as the standard. Broadly, however, it works either by an individual or group coming to the school or by work samples and/or assessments being taken to a central or regional moderating group.

There are many permutations, including:

- a teacher simply observing how others assess and being left free to make his/her judgement,
- collective judgement — say, teachers from a region assessing work samples from a class,
- a 'king' moderator, who has internalised a standard and inspects classes in order to give the teacher assessment his imprimatur.



CATCH 22: *If the marks don't correlate well, it's not legitimate to add them up; if they do correlate well there's no point in having teacher assessment.*

At the very least, any consensus model would involve discussion amongst teachers within a subject about curriculum and assessment. The existence of HSC as an external exam system has meant that teachers have been left largely in the dark on technical matters relating to assessment. Consensus moderation, if implemented seriously, is a process that could begin to put assessment back in the hands of the classroom teacher. However, to a conservative politician, it hasn't much going for it. It will mean greater expenditure on staffing, it will mean less centralised control of the curriculum, it will mean greater understand-

ing of the whole process by teachers.

Teachers must have additional meetings in order for moderation to take place. Time release is therefore a basic condition for its operation.

What is holding back consensus moderation?

The VISE policy statement was announced in April 1978. In relation to consensus moderation it stated, "Experimentation and pilot studies in different subject areas will be needed."

Two years further on, members of VISE Committees asking about vital questions of time release for teachers involved in consensus moderation schemes were told that VISE had not yet taken the matter up with the Government. The first meeting took place in May of this year, when VISE was informed that no definite answer could be given whilst the Minister was overseas. Yet another case of bad timing?

One thing is sure: consensus moderation will involve additional staffing. Whether the Government makes the necessary budget entry under VISE or the Education Department is neither here nor there. The key question must be: "Why hasn't a financial decision of such significance been raised before?" And further: "Is VISE likely to campaign more aggressively than Secondary Division for funds?"

Group 1 English:

It was in this context that the HSC English subject committee last year lodged their course for VISE Group 1 Accreditation. The submission included a request for consensus moderation, with details, plans and budgets for the necessary in-service programs.

Many months later, after considerable to-ing and fro-ing, the proposal was endorsed by the various VISE Committees—Assessment Procedures, Accreditation and Curriculum and Assessment. At its July meeting, VISE Council decided that statistical moderation would be used.



It is at this level that political priorities and questions of control start to have major implications for the implementation of reforms in education as set out in the VISE Policy Statement. One particularly influential committee in VISE is the Administration and Finance Committee. One particularly influential member of this Committee is Mr T. J. Ford. Mr Ford makes no secret of his opposition to consensus moderation.

The Minister, whether at home or abroad, has enough close contacts in VISE to know whether to meet them or not. Questions of extra staffing and of curriculum control slipping away from the centre are better left unasked, particularly when policy statements state the educational necessity for devolution.

A number of questions are raised, the most important being: Can a subject accredited with one set of assessment procedures legitimately proceed when those methods will not be implemented?

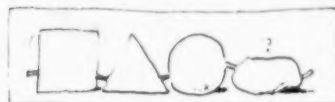
The options devised by the English Subject Committees—Oral English, Creative Writing and Focus on Language—cannot validly be moderated against the exam.

The VISE Council decision does have a positive aspect to it, however: English teachers have been shown that the way towards real development in their subject is not within Group 1. They would be well advised to watch very closely for the development of Group 2 English courses which are not subject to the same constraints.

*It just could be that
English is the point at
which the aggregate
grip on HSC can be
broken . . .*

The fact that Group 2 subjects are entirely school-based means that statistical moderation cannot be imposed on them. Nor can standardisation. It just could be that English is the point at which the aggregate grip on HSC subjects can be broken. Sooner or later the number of subjects that cannot be standardised, statistically moderated and homogenised down to a single numerical mark will be sufficiently large to render this absurd rank ordering selection method inoperable. If enough of us put our minds to it, it could be very soon indeed.

Phil Nouse is a member of the VISE Assessment Procedures and School-based Curriculum Development Committee.



Appendix 7

From *The Secondary Teacher* No. 2, 1980

STANDARDISATION — what is it? Where did it come from? Why is it there? What is its rationale, what are its assumptions? Any thoroughgoing reform of the Victorian Higher School Certificate cannot even be contemplated without these and many related questions being broadly understood by secondary teachers. For fifteen years in one form or another it has dominated the HSC,

virtually unquestioned, totally removed from the experience of the classroom teacher.

How does it fare when its operations are exposed to analysis? Could it continue in the face of increasing teacher opposition? What are the alternatives? The 'Secondary Teacher' invites speculation and information on an issue that is crucial to the directions VISE will take in the next few years.

Where some
are more equal
than others.

STANDARDISATION

UNIVERSITY SELECTION is too important to be left to academics — it's all done by clerks.' So mused last year a member of the VISE Assessment Procedures Committee. Selection methods based simply on adding the applicants' HSC scores and placing them in rank order have the big advantage of being cheap in terms of tertiary budgets and staffing. They can be passed off as objective and concise, and hence be used as alibis against any accusation of discrimination. That the method's attributes don't run to any predictive validity in that area of study is widely acknowledged, a weakness which has in most tertiary faculties been traded off against the above bonuses. Alternatives cost more and generally involve subjective, personally accountable decisions.

Standardisation is the process that legitimises aggregating marks. If all subjects can be sensibly compared one against the other, and if a 50 in Physics can be made to be 'worth' the same as a 50 in History, and so on, a very useful indicator of general ability can be established. It's on these very iffy 'ifs' that the entire HSC assessment system teeters, as it has in its present form, since 1971.

To paraphrase the above quip about selection methods, standardisation derives its existence from the notion that satisfactory performance in a subject is too important a decision to be made by the subject examiners — it's all done by the computer mulling over the candidate's other marks.

Less flippantly, a choice has to be made about how cut-off points in a ranking order are to be made. The problem with leaving that to the examiners is that subjects will produce a range of pass rates, with no justification say, for a 60% pass rate in Maths against a 75% pass rate in History. History could be termed a 'soft' subject. Worse still, this lack of comparability makes aggregating marks meaningless. The standardisation method would, theoretically, change all that.

The first crude method, employed from 1964 to 1971, simply defined 67.8% as the target pass mark for all subjects. Marks

By **PHIL NOYCE**,
Member of the VISE
Standardisation Sub-committee

were added up, University quotas were filled with the top scorers. University professors at this point stopped making the decisions about the HSC pass rate and the computer took over.

The 67.8%, incidentally, was derived from the average of all Matriculation subject pass rates in 1960, 1961 and 1962. It turned a blind eye to questions of whether one subject may actually be of a more difficult nature than another or that bright students might choose 'hard' subjects and so on. The new orthodoxy simply asserted: All subjects are now equally difficult to pass. Students are spread evenly in ability.

The great homogenising process rolled on. It seems almost a digression to suggest that, by their nature, subjects are not all equally anything. There are demanding subjects, satisfying subjects, technical subjects, relaxing, informative, useful subjects, etc., etc. Wrong — they are all equally difficult because, if they're not, they can't be added up.

By the late sixties, researchers started putting the 'difficulty' assumption under the microscope. But what microscope? Teachers' experience? Examiners' opinions?

Hardly objective enough to dislodge the enshrined 67.8%, which was there because it was there.

Hence, a search for other measuring devices . . .

Most fourth form students sat for the Commonwealth Scholarship exams. Results from that, combined with comparison of random groups of students' performances on different subjects indicated that a good deal of tuning was necessary if standardisation was to be maintained. Modifications were made, which led to the present form, which has operated since 1971.

THE CURRENT PRINCIPLE:

For any subject, the distribution of standardised marks of the candidates in that subject should agree with the aggregate distribution of standardised marks in the other subjects taken by these same candidates.

More sophisticated than 'everything's the same', and carrying a few more assumptions, to wit:

Assumption 1:

● The 'ability' of candidates in a particular subject is determined by their performance in their other subjects.

Thus:

- 78% of Chemistry students obtain a D or better precisely because 78% of them got a D or better in their other subjects.
- 64% of Australian History students obtain a D or better because that's how many of them passed the other subjects they did.

Assumption 2:

● Average 'ability' across all subjects is constant from one year to another.

This is necessary in order that a total subject distribution can be constructed to have a mean of 57%, a standard deviation of 15 and a 'D' cut-off point of 68% of candidates — figures chosen purely because that's

ADVICE TO A STUDENT WISHING TO GET ON:

I WANT TO DO LAW AT UNI.



DO HSC MATHS & SCIENCE.

Professors may later condemn your ignorance of history, language and culture. They may roll at a system that provides them with a student with only one demonstrable skill — passing exams. But you will be on.

I WANT TO DO MEDICINE AT UNI.



DO HSC MATHS & SCIENCE.

Take care not to do Biology, or Environmental or Social Sciences. The bad company you keep will depress your score. On the question of company, choose parents who are doctors. That correlates with successful entry better than any known criterion.

I WANT TO DO HUMANITIES.



DO HSC MATHS & SCIENCE.

It's those scores again. Actually, it's not so crucial here — apart from Melbourne Uni. you can practically walk in to whatever you like.

I DON'T WANT TO GO TO UNI, OR DO ANY TERTIARY STUDY — BUT I WANT MY HSC.



DO HSC MATHS & SCIENCE.

With a percentage pass rate about 12% or more above all the others, you increase your odds. Besides, employers believe in it.

what happened to be the case in 1970, the year before the present system was introduced.

Assumption 3:

● The Normal Distribution is an appropriate model for spreading marks in all subjects of all candidates.

That is, the assessment system assumes an overall random spread of results.

Before examining these assumptions, let's look at the fundamental decision which renders them necessary, valid or not. It is this: that examiners' opinions in subjects with over 500 'normal' candidates can have no bearing on questions of:

- satisfactory performance
- distribution of As, Bs, Cs, Ds, etc.
- special distinctions
- pass-rates.

Supported on the above assumptions and armed with this decision, the computer churns out the "auto-standardised" spread of marks and pass-rates for the 22 subjects with over 500 N-type candidates (those doing at least three subjects plus English).

For the 33 subjects with less than 500 N-type candidates the procedure involves a compromise between auto-standardisation and the examiner's suggestion, weighted accordingly. For example:

For 100 candidates: weighting is 4 to 1 in favour of examiner's suggestion;

For 400 candidates: weighting is 4 to 1 in favour of the computer.

So conditioned are we to the HSC running as it is that we tend not to think about what it means to have pass-rates determined in the absence of an examiner's opinion. To give it some perspective, let's imagine the outrage that would result if either of the following situations were suggested:

1. The administrations of each of the Universities inform all the faculties that, in the interests of uniformity, all pass rates be determined centrally, and without regard to examiners' opinions or faculty objectives. Preposterous! Yet it was the universities which required that to be done at the introduction of the HSC and the demise of Matriculation in 1964.
2. All fifth form teachers set, mark and order the set of end-of-year exam papers. Teachers are instructed not to make any value judgements about what constitutes a satisfactory result, or a balanced distribution of marks, but to leave these considerations to a Department official who would choose the paper below which performance was considered unsatisfactory. This



"THE 'ABILITY' BEING MEASURED IS THAT QUALITY WHICH PRODUCES GOOD RESPONSES IN A 3-HOUR EXAM."

person would also make decisions about the quality of the top result and the spread of As, Bs, Cs, etc.

How would such a proposal be received? It surely sounds outrageous — inspection in its heyday was never so demeaning in attitude to teacher competence: inspectors might wish to see a marked paper to gauge your judgement, but to divorce the teacher entirely from the evaluation process was, even in those dark ages, unthinkable.

What happens with HSC? As we well know, classroom teachers have nothing to do with marking their students' papers, the examiners carry out instructions according to the above limitations and the computer strikes the pass mark.

It is true that some HSC subjects presently do, and from 1981 all Group 1 subjects will, include a moderated, teacher-assessed component. Questions related to moderation can't be gone into in this article; suffice it to say that the hierarchy of examiner, computer in these cases is extended to teacher, examiner, computer.

By contrast, the above fifth form scenario, repugnant as it is, appears relatively humane — the classroom teacher being linked to, if not in control of, the assessment process.

I labour the point to emphasise this fact: that there are no plans to date to alter the standardisation process for Group 1 subjects, nor have any conceivable improvements to it been expounded.

In my view, standardisation is no more than a device to carry out the requirements of university administrations. It cannot be simply removed without a total review of tertiary selection procedures. Any readjustment of it is likely to produce other anomalies.

Given the nature of the task to be performed, it probably works as well as anything yet devised. Critics of standardisation therefore need to examine not so much its procedures, but the rationale on which it rests. Do away with the need for aggregating marks, and standardisation would be redundant.

To date, the most trenchant criticism has come, not surprisingly from those discip-

Subject	1979 Pass Rate
Chemistry	78%
Physics	78%
Applied Maths	77%
Pure Maths	75%
General Maths	75%
Subject Average	68%
Economics	66%
Geography	65%
Politics	65%
Australian History	64%
Accounting	64%

lines on the losing end of the system—Italian, Greek, Graphic Communication and Australian History, to cite the most vocal. As can be seen from the above selected table of pass rates the straight Maths/Science subjects emerge as clear leaders for big subjects in the pass rate stakes.

Why is it so? The computer says, based on performance in their other subjects, science students are a 'high ability' group, which doesn't really answer the question.

The 'ability' being measured is that quality which produces good responses in an end of year 3-hour exam in their other subjects. The simple observation, endorsed by the computer and defined as a measure of ability, is that if candidates perform well in Pure Mathematics, it's a good bet they'll do well in Applied Maths, Physics and Chemistry. It just works that way much better than, say, in History, Politics and Economics.

How is the physical science/social science differential maintained?

Roughly, by the computer picking out those students doing say, Physics and History, for example. They tend to do relatively poorly in the Physics exam compared with those who do Physics and Chemistry. They debate the History currency, and accordingly, its pass mark has to be made higher and harder.

The real reason?

Mark true or false (T or F) for the following:

	True	False
Physical science candidates are basically more 'intelligent'	☐	☐
Physical science examinations are relatively similar in style, and test similar skills	☐	☐
Examinations are best suited to objective questions and convergent responses	☐	☐

If you scored F, T, T then I agree with you. The current system tests a very narrow, specific form of ability—that of performing well in a high anxiety, artificial situation requiring convergent, rapid-fire responses. Undoubtedly it is a form of 'ability' being measured. But what of other skills and qualities—of tenacity, of perception, of consistency, of interest, of non-linear ways of thinking, of the myriad other human qualities which, in the execution of a particular job, become far more relevant than that which HSC tests directly.

The question hinges very largely on the notion of 'intelligence', a term largely appropriated by the intelligence testers, ranging from the IQ nuts to the disciples of Bloom's Taxonomy. Broadly, their position has the following practical result:

What can't be measured can't be tested.
What can't be tested can't be assessed.

What can't be assessed is relegated to the curriculum graveyard.

I, of course, have no empirical evidence to suggest that social science students (and teachers) are "more able" than those who populate the Physics lab. But then I'm no more able to decide on whether $E = mc^2$ or Hamlet is the "more important" contribution to western civilisation.

For non-Science/Maths teachers there would seem to be two broad avenues for action:

1. Within Group I—Bunch subjects with identical or similar exam formats, eliminate extended response, discursive questions requiring opinion and analysis and instead construct questions with a correct answer.—'Ability' as measured on 'It's Academic', 'Mastermind', and the like. It may not extend the other sorts of ability you are presently developing, but it's likely to bring you closer to the science results.
2. Group II takes you out of the standardisation game and into considerably greater control over teaching to your interests and to student use. It also opens up the possibility of assessing a much wider concept of 'ability'. As accreditation procedures develop in the Group II area, we can expect to see increasingly diverse assessment methods which more broadly reflect overall performance.

Let's now take a look at the main assumptions underlying standardisation, using some selected results from the 1979 HSC scores:

Assumption 1:

● Candidates' 'ability' in a subject is determined by performance in their other subjects.

In most cases this might seem a fair statement. Candidates in Pure Maths are likely to be 'equally able' in that subject as they are in Applied Maths, perhaps also

in Physics. It might be stretching it somewhat too far to assume that these candidates also proved 'equally able' in, say, French.

Languages provide a good area for testing this assumption for three reasons:

- It is not at all self-evident that having a talent for languages in general would mean that a person would therefore perform well in History, Art, Physics or whatever.

- Demographic and socio-economic factors related to Australian immigration further complicate questions of general ability in languages. For example, it's unlikely that kids in Melbourne who speak Italian would perform as well even in a closely related language such as French as they do in Italian.

- Languages are generally small subjects. With the exception of French, all examiners' suggestions are included on the computer print-out, thus providing ready comparison between estimates of satisfactory performance as determined by computer standardisation and examiners' opinion.

A brief look at results in a few languages might give an indication of the problems involved.

Hebrew has a number of candidates who do it as a compulsory subject, at schools such as Mt. Scopus. The examiner is concerned that it is not being treated seriously enough by its students who instead devote their energies to their other subjects. We would expect that if this attitude was widespread, those students would not deserve excellent results in that subject. In fact, in the examiner's opinion, only 61% of them deserved to pass.

Auto-standardisation would reward those students, in Hebrew, for not concentrating on it. Instead of the examiner's 50, the pass mark would have been set at 13 if the computer carried out its orders on auto-standardisation. As it's a small subject, an interpolated compromise between 13 and 50 had to be arrived at. The computer compromise for the D cut-off point in this case turned out to be 39, which was endorsed by the Standardisation Committee.

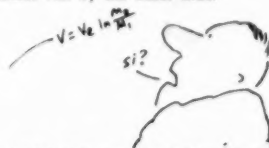
In percentage pass terms, the compromise between the auto-standardised 88% and the examiner's 61% was 74% for Hebrew students.

For practically all other language subjects the computer compromise between examiner's suggestion and auto-standardisation was seen by the Standardisation Committee to be too harsh. Thus Chinese, Dutch, Hungarian, Indonesian, Japanese, Latin, Modern Greek, Russian, Serbo-Croatian, Spanish, Turkish and Ukrainian, all had the D cut-off point lowered to allow a higher pass-rate more closely resembling the examiner's suggestions.

The figures for the largest two groups of non-English background — Greek and Italian — illustrate the systematic penalising of students who do have a special ability in particular subjects. Because Italian has only recently reached 500, the examiners' suggestion was included on the print-out, although no automatic compromise could be made, given the rule that all subjects over 500 candidates operate under auto-standardisation.

Out of the 514 N type candidates, the examiners considered 428 had demonstrated a level of competence that merited a pass. The computer, picking up the data that they weren't that good in Physics, English, Australian History and whatever else those 534 candidates did, would have decided to allow 319 through — chopping 109 students off the examiner's reco.

In the discussion on Standardisation Day that ensued to settle on a percentage pass rate somewhere between the computer's 'pass' rate of 59.74% and the examiner's 80.15%, I threw in a personal anecdote relating to having taught students in Italian, Mathematics, Music and Science. It was not really the most provocative of assertions — that I had taught kids who were whizzes at Italian and drongoes at Maths — they tended to have names like Sandra, Franca, Pasquale and Maria — whereas Stephen and Zoran, who could factorise a quadratic as quick as look as you, ended up getting remedial Italian tuition in special classes run by the other kids.



"THE ABILITY OF CANDIDATES IN A PARTICULAR SUBJECT IS DETERMINED BY THEIR PERFORMANCE IN OTHER SUBJECTS."

Subject	Number of 'N' Candidates	Auto-standardisation	Examiner's Suggestion	Computer Compromise	Decision
Japanese	75	85%	100%	99%	100%
French	1,207	81%	—	—	81%
Hebrew	144	88%	61%	74%	74%
Greek	454	52%	75%	55%	67%
Italian	534	60%	80%	—	66%
Serbo-Croatian	46	41%	98%	98%	98%

Eventually, we decided on 66% because, well, because . . . The same story for Modern Greek: auto-standardisation's 52% and examiner's 75% producing the computer compromise of 55%. Something had to be done, given the Italian decision, the Greek examiner's paper warfare on the matter in recent years, and the quite untenable assertions that the Greek candidates, all but one of whom have Greek parents, should be expected to do no better at Greek than they do in English, Australian History, etc.

A tolerance mark comes with the print-out, by which amount we can justifiably change the result—for Greek at the D level it's 4.7%. Adding that to the 55% gets close to 60%. We decided on 67.6%. Why? Because, well, because . . .

And if the trends in the graph at right are any guide, things aren't going to improve, particularly if it's a growth subject. One way of Modern Greek improving its results would be to cut back on its numbers, fiddle the examiner's recommendation in

anticipation of standardisation cutbacks, take a leaf out of the book of Hebrew and instruct students to devote more time to the other subjects. It's a fiddle I'd seriously propose to those wishing to fight the computer.

Fiddling the marks is hardly a novel suggestion. The measurement heavies allege that it has been going on since standardisation began, indicative of the fact that the examiners can't be trusted. The argument is usually taken one step further to justify the essentially political bases on which standardisation rests:

(1) Statistical methods are necessary in order to ensure comparability between subjects, because the examiners can't be trusted.

(2) If external, impersonal examiners can't be trusted not to fiddle the results, unbridled classroom teacher assessment would be tantamount to anarchy.

That research carried out in this State and others at Year 12 fails to substantiate

THEY SAID:



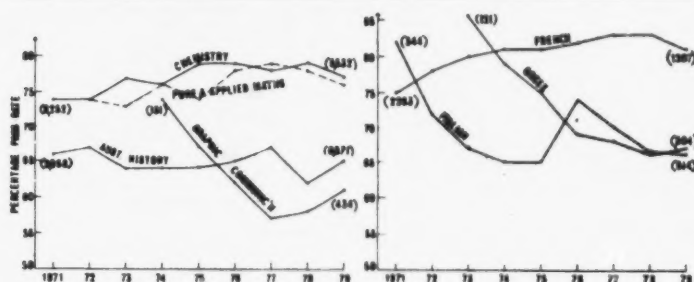
"REGRETTABLY, the Victorian universities have been but lukewarm in their reaction to the VISE initiatives. They have shown a dogged determination to stick to the HSC score formula as the main basis for university entrance. I FIND THIS SYSTEM ONE OF THE MOST INHUMAN IN THE WORLD. One point can mean the difference between admission or rejection, between acceptance for your chosen discipline and an unacceptable alternative. And this with no account being taken of the type of school attended, the standard of teaching, the home background or the motivation of the applicant. Now this might be defensible if there was overwhelming evidence that this method of selection was successful. BUT MOST OF THE EVIDENCE IS TO THE CONTRARY. Only in the Physical Sciences and Mathematics is there a high correlation between HSC score and program at university. In the Humanities, the correlation is virtually nil. All universities have in recent years taken in an increasing number of mature age students with few formal qualifications but highly motivated. These students have been enjoyable to teach and have done well. 99

—FROM AN ADDRESS BY PROFESSOR SCOTT, VICE-CHANCELLOR, LA TROBE UNIVERSITY, JULY, 1979.

"I have always been of the view that the proper preparation at school for studying law at the university is a humanities course, including such subjects as history and Latin. As long, however, as the Law faculties at Melbourne and Monash universities require as the only qualification for entry a minimum score in Higher School Certificate of better than four 'As' (Melbourne) and better than two 'As' and two 'Bs' (Monash), so long will the candidates at school (not necessarily their teachers) think in terms of scoring the maximum number of marks in the subjects in which they think they can do best, which in very many cases means presenting in mathematics and the sciences, rather than in humanities subjects. 99

—N. A. H. CREESE, HEADMASTER, MELBOURNE CHURCH OF ENGLAND GRAMMAR SCHOOL, IN "THE AGE", FEB. 8, 1980.

How 'pass'* rates in some selected subjects have fared under standardisation from 1971-1979



*The term 'pass', as used here and throughout this article, whilst generally understood, has no technical meaning in VISE terminology. It is used here, as elsewhere, to describe the 'D' cut-off point, or the standardised mark of 50 for any subject.

() The numbers in brackets indicate the number of candidates in that subject in that year. Generally, figures for 1971 and 1979 have been used, except for subjects which started after 1971.

In all cases, growth or decline in the period has been relatively uniform — e.g. Greek, Italian and Graphic Communication have grown steadily, French has declined steadily. Note the corresponding changes in 'pass' rates.

these attitudes as well as a general lack of concern about it at all in years below Year 12, has not diminished the ardour with which this quite reactionary view is held by key people at top policy levels.

* * *

There are however, more fundamental, educationally valid, practical solutions. Without even getting into analyses of the vested interests of that small number of elite pockets such as Melbourne University's Medical Faculty which make standardisation a political reality, by-passes have already been created. To date, the current HSC subject committees have decided to continue travelling along the known road — to be renamed Group I from 1981 on. Many of them write letters to VISE complaining how rough the road is and how many casualties it sustains by comparison with their recommendations.

Group II does, among other things, provide a way around standardisation. Computers can't be programmed to invent a pass mark when a subject doesn't have numerical ranking. Satisfactory performance would have to be decided in some other, non-mechanical way. An assessment in Greek, for example, would be totally independent of that student's assessment in English.

All of which is known to those who plan to stay on Group I and whimper

'we'd rather stay with the strength' each time they get run off the road by Physics, Maths, etc.

Whilst another road has been created, older travellers seem to be waiting for newcomers to try it out.

* * *

If standardisation were relevant to those small subject pockets, there might be a case for continued submission. But it's not as if the Chairman of the Panel of Examiners in Modern Greek has to convince the Modern Greek faculty at Melbourne University that his assessment of student ability to do further Greek is reliable — he sits in both places. By a similar process in other subjects, a student could conceivably be assessed as suitable to enrol in an Arts Course at Melbourne University. Going a bit further, the three hour exam could be replaced with interviews, folios, projects and other techniques as practised by, among others, the STC group of schools — methods being accepted by a growing number of tertiary institutions. And remember — of the 55 HSC subjects, 33 in 1979 had fewer than 500 candidates.

If I'm harping on university entrance, a question of marginal concern even to HSC students, I'm doing so because this system designed for university selection, has been used by employers and others for a variety of uses and abuses, by default. That it has been condemned by everyone to the left of Malcolm Fraser has to date done little to change it.

Assumption 2:

• Average ability across all subjects is constant from one year to another.

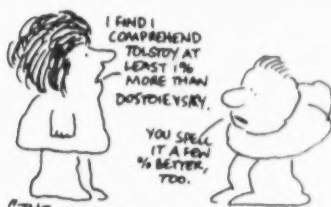
Turn that around a bit. If some subjects increase their pass rates, others must suffer a decrease. In statistical parlance, French, for example, has attracted students of a progressively higher 'ability' group, rising steadily as it has from 1971 with a pass rate of 75% to 81% in 1979. In that time its numbers have fallen steadily from 2416 to 1253. Greek, on the other hand, a 'growth' subject, going from 121 to 511 over the past seven years, has seen its pass mark plummet steadily from 86% to 67%.

And anyhow, what basis is there for asserting that HSC kids in 1979 are as 'able' as was the class of '69, or as those in '89 will be? Simply, that without it, we couldn't have standardisation.

Assumption 3:

• "The normal distribution is an appropriate model for spreading marks."

The normal distribution tells you the spread of results you get from a sample of random events. It predicts a 75% chance of not getting five heads turn up in 10 tosses of a coin; it can provide estimates of how many people on a given day will walk into Myers wanting a shirt to fit an 18" neck.



"THE LOGIC OF THE COMPUTER... ALLOWS STATEMENTS LIKE 'THERE HAS BEEN A 1% INCREASE IN ABILITY IN THIS SUBJECT.'"

Certainly, for examination results, statistical evidence can be used to justify the use of the normal distribution instead of people's opinions for solving questions of pass and fail. But it does other things as well. Examiners fudge. They start bolstering a mark here, fattening a kink there, in anticipation of the Great Leveller. Further down the line, those without recourse to manipulation, the teachers, see the system as one entirely out of their control.

It has acted to repel many teachers from Year 12 who find the educational implications repulsive and to attract at least an equal and opposite number who take solace in it. Moreover, education is surely a purposeful, directed activity; learning hopefully taking place more than by sheer chance.

Whereas people tend not to aim for an 18" neck, or generally not toss double-headed coins, teachers and students do try to influence the result. If any distribution could be chosen to fit such an enterprise, it would surely be skewed in favour of the 'success' end. I'm not advocating a change-over to the Poisson as a better distributor of results; working from the basis that no matter what your efforts are, the results will fit any predetermined shape is, in my view, not a sound philosophical basis for education.

Graphic Communication illustrates the point well. According to the examiners, the subject has undergone considerable course development, teachers are more qualified and experienced in a period when the number of candidates has risen from 151 in 1974 to 409 in 1979. Why have its pass rates plummeted from 74% in 1974 to a low-point of 57% in 1977? Quite simply because more are doing it.

Whereas standardisation has consistently recorded around 50% for Graphic Communications the examiners have usually suggested around 80%. As more kids take on this subject, a group even by computer logic of much the same 'ability' level, the pass rate goes steadily down. This year the interpolated pass rate was 54%. Very sloppy correspondence from the examiners certainly influenced a Standardisation Subcommittee decision to override normal practices and boost this figure to 61%.

Hard statistics superimposed on subjective opinion — in engineering terms, the equivalent of bitumenising quicksand.

'Tolerance' as a statistical concept is based on the premise that there is accuracy in numbers. Imposed as it is on a very wobbly foundation of the divergent, subjective opinions of examiners, it provides a veneer of accuracy to the total assessment procedure. An engineering equivalent to this application of statistics might be found in the process of bitumenising quicksand.

In the case of English with 22,465 candidates (incidentally, the lowest since 1971), the computer's estimate of the pass mark is accompanied by a very small tolerance factor — 0.6 of a mark in fact, compared with 1.5 marks for Applied Maths, and 14 marks for Serbo-Croatian. It's got nothing to do with the nature, or the degree of subjectivity involved in the assessment of discipline.

Countless case studies have demonstrated yawning variations—dozens of marks between examiners, and even by the same examiner at different times. Not surprisingly these variations between examiners are greatest in subjects like English. The logic of the computer is entirely different, allowing statements such as "there has been a 1% increase in ability in this subject" to be uttered as a commonplace statistical shorthand by people fully aware of what a nonsensical statement this is in real terms.

As it turned out in the case of English, questions more of a political nature overrode the .6 of a mark tolerance. The auto-standardisation pass mark of 50 produced a pass rate of 68%. As English is a compulsory subject, it is treated as a special case for variation of the auto-standardised score. A study of the cut-off point revealed the following: Making 49 the cut-off point would allow another 731 students through, 48:1436, 47:2013 and so on. Where to draw the line? An examiner's opinion? Certainly not. Such an approach is totally in contradiction to the very principle of standardisation. It comes down to numbers—and administrative ease. One telling line concluded the discussion: "48 would save us the headache of receiving letters from 1436 angry people." And that was where we stopped.

From all this I conclude the following:

Next year, present HSC subjects will generally be Group 1 subjects unless the relevant subject committees indulge in some very quick footwork. All the evidence suggests that if they wish to get good pass rates, they would be wise to follow this advice:

1. Keep the numbers down. For large subjects, culling out the ones who are not doing Maths and Science is possibly the clearest criterion. If you must do a History, don't pick Australian—it's too popular. At 64%

pass rate, it doesn't compare with Eighteenth-century British History's 77%.

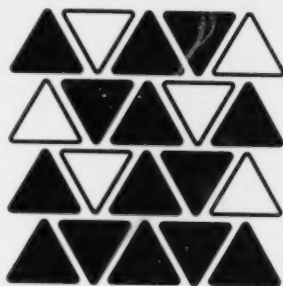
2. For subjects under 500, the smaller you go, the more power to the examiner's arm. Serbo-Croatian, for example, if it shows signs of growth, will end up with pass rates looking increasingly more like the computer's 41% and less like the examiner's 98%.
3. Encourage your students to choose their other subjects in terms of high pass rates and then work hard in them, in order to bolster the pass rate in your subject.
4. If the above measures are not to your taste, get out of Group 1.

Footnotes:

1. For a more technical explanation of the standardisation process, see the paper prepared by VISE Liaison Officer Murray Cropley and published in VISE Circular No. 4, 1979.
2. I wish to acknowledge the assistance of VISE Liaison Officers in their advice on factual matters related to the preparation of this article.
3. The Standardisation Sub-committee met on 'Standardisation Day', Sunday, January 6, 1980. This meeting made all related decisions for HSC subjects. Voting rights applied to those members of the VISE Assessment Procedures Committee who were in attendance, one of whom was the author of this article.
4. This article barely touches on a whole range of related issues. 'The Secondary Teacher' wishes to encourage discussion on a topic which has for too long been a total mystery to teachers. Contributions, reactions and views of any sort on the matter will be eagerly received.

Source: Victorian Secondary Teachers Association, Year 12, VSTA, Richmond, Vic., 1984, pp. 4-17, 21-30.

Annotated bibliography



Blachford, K. 'To B or not to B'. *Secondary Teacher*, no. 8, October 1981, pp. 18-20

A concise account of the argument against grading from a Victorian point of view.

Block, N. S., & Dworkin, G. (eds). *The IQ Controversy*. Pantheon, New York, 1976.

Reprints of the last half-century's debate on the basis of intelligence — genetic or acquired?

Dow, G. M. 'Streaming'. *Secondary Teacher*, no. 157 July 1970, pp. 21-8

Although it was written in 1968, this still stands as a fair summary of the research on the value of streaming and other forms of grouping.

Hannan, B. *The Democratic Curriculum*. George Allen & Unwin, Sydney, in press.

A collection of writings, mostly from Victorian teachers' journals, from 1968 to 1983. Grading and credentialling arguments occupy at least half of the collection. Refers specifically to the debate in Victoria.

Hoffman, B. *The Tyranny of Testing*. Crowell-Collier, New York, 1962.

A lively attack on the assumptions behind mass testing, particularly in the American context.

Houts, P. (ed.). *The Myth of Measurability*. Hart, New York, 1977.

An American collection of writings attacking standardised testing and IQ.

Kirchenbaum et al. *Wadjet?: The Grading Game in American Education*. Hart, New York, 1971.

An attack by various American writers on the educational, or anti-educational, effects of grading and credentialling practices.

Rosenthal, R., & Jacobson, L. *Pygmalion in the Classroom*. Holt, Rinehart and Winston, New York, 1968.

A classic experiment in the self-fulfilling nature of teachers' expectations of students. Although it has been much disputed, it played an important part in discrediting streaming.

Rowntree, D. *Assessing Students: How Shall We Know Them?*. Harper & Row, London, 1977.

The most illuminating general account of current assessment issues and practices.

Simon, B. *Intelligence, Psychology and Education*. Lawrence and Wishart, London, 1977.

An account of the role played by psychological theory in the development of education, especially in England. Examines questions such as cultural bias in testing.

Simon, S., & Bellanca, J. A. (eds). *Degrading the Grading Myths: A Primer of Alternatives to Grades and Marks*. Association for Supervision and Curriculum Development, Washington, DC, 1976.

An American collection attacking grading and proposing other methods of assessment.

About the author

Bill Hannan spent the first twenty years of his working life teaching English and modern languages in Victorian secondary schools. Between 1967 and 1972 he worked with the Curriculum Advisory Board both as an executive officer and as one of the teachers implementing curricular change, first at Moreland High School and subsequently at Sydney Road Community School. From 1963 he was the editor of the Victorian Secondary Teachers Association's magazine *The Secondary Teacher* and a member of the Association's executive. He left teachers' publications in 1973 to work as a writer and a theatrical director, and returned to them in 1980. He is now an editor of *The Australian Teacher* and *The Victorian Teacher*, and a member of the Council of the Curriculum Development Centre.



Acknowledgements

Reading 1: Reproduced by permission of Patricia Broadfoot, the Falmer Press and the author; Reading 2: Reproduced by permission of Melbourne University Press; Reading 3: Reproduced by permission of the Australian Association for the Teaching of English and the author; Reading 4: Reproduced by permission of the Victorian Institute of Secondary Education; Reading 5: Reproduced by permission of the Victorian Secondary Teachers Association.

Deakin University



ESA844 Administrative Context of Schooling
Assessment and Evaluation in Schooling
ISBN 0 7300 0300 0
Cat. no. ESA844M03